

11 June 2010 | \$10

Science



Helga Nowotny is president of the ERC and professor emerita of ETH Zurich. She is chair of the ESOF 2010 Scientific Programme Committee. Her research focus is social studies of science and the coevolution of science and society. E-mail: helga.nowotny@wwtf.at

Frontier Research in Europe

NEXT MONTH, THE FOURTH EUROSCEINCE OPEN FORUM (ESOF 2010) OPENS IN TORINO, ITALY, TO discuss the frontiers of science and technology in Europe and to stimulate policies that bridge the gap between science and society. In this context, the current European research landscape can hardly be imagined without the innovative funding programs of the European Research Council (ERC), which has selected its 1000th grantee just over 3 years after its launch. As the new president of the ERC, I am particularly pleased to continue building on past achievements while preparing for the challenges ahead.

Established by the European Commission (EC) in 2007 as part of the Seventh Framework Programme of the European Union (EU), the ERC funds curiosity-driven bottom-up research by individual investigators based exclusively on the principle of scientific excellence in the 27 EU member states and “associated countries.” The ERC is supporting a new generation of young researchers who seek to break out of academic hierarchies and their national funding systems to obtain early scientific independence. It also supports established researchers in pursuing risky ideas that might lead to new discoveries across all scientific disciplines. But beyond providing trustworthy and fair funding opportunities that are open to researchers from all over the world (provided that they work in Europe), the ERC carries European added value. For the first time ever, universities in Europe compete for the prestige of hosting ERC grantees, acknowledging the importance of internationalization and of nurturing the most talented younger researchers. In the new EU member states (such as Poland), serious efforts are being made to revamp national granting systems, following the ERC model. Moreover, a number of countries have decided to rely completely on ERC evaluation by providing research support to semifinalists of the Starting Grant Call who could not receive ERC funding due to budgetary constraints.

But questions persist about the institutional structure of the ERC. An unusually frank report by a Mid-Term Review Committee in July 2009 laid bare structural problems related to its governance and procedural issues, such as the registration of remote referees and reimbursement of panel members. The recommendations were reinforced by the Competitiveness Council of Member States in March 2010, which endorsed another review before July 2011 to assess whether structural adaptations are sufficient or a new institutional model must be set up. At the heart of the matter is the dual construction of the ERC. Its Scientific Council, consisting of 22 eminent scientists, sets the scientific strategy, whereas the administration and implementation are in the hands of an Executive Agency, composed of nearly 300 staff (with about 45 Ph.D.s). The question is how to better align an ambitious and vibrant world-class funding agency with the rules and regulations of the EC. The recruitment of a distinguished scientist with robust administrative experience as the new director of the Executive Agency is now under way, merging the present functions of the ERC secretary-general and the director of the Executive Agency administrative side.

The future of the ERC beyond 2013 depends on three issues. It must become a permanent institution *sui generis*, with a structure and mechanisms adapted to its unique mission. Its annual budget must be at least twice its 2013 budget of 1.7 billion euros to ensure sustainability. And while continuing to fund individual excellence through its bottom-up approach, the ERC must expand its program in a forward-looking, innovative way.

At a time when European policy-makers seek to bolster innovation to meet the grand challenges ahead, Europe must vigorously pursue frontier research as the excellent science base from which many new discoveries will arise. The forthcoming ESOF 2010 conference is an important forum for discussing how the future of the most exciting European funding agency can help Europe prosper.

– Helga Nowotny

10.1126/science.1190995



Downloaded from www.sciencemag.org on June 10, 2010



Suresh nominated
to head NSF

1340



The roots of
Jewishness

1342

SCIENCE AND THE LAW

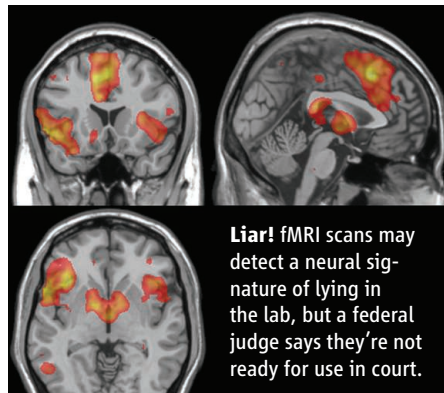
fMRI Lie Detection Fails a Legal Test

Lie-detection technology that uses functional magnetic resonance imaging (fMRI) scans of brain activity is not yet ready for use in the courtroom, a federal magistrate judge in Tennessee decided last week. Defense attorneys for Lorne Semrau, a psychologist accused of defrauding Medicare and other health-benefit providers, had sought to introduce fMRI brain scans to show that their client had no intention of cheating the government and other insurers. But after a pretrial hearing in mid-May that featured testimony from scientists on both sides of the issue, Magistrate Judge Tu Pham concluded that the scans don't measure up to the federal courts' standards for scientific evidence. Pham's 39-page report is the most formal legal opinion yet on this controversial technology, and although other courts don't have to follow suit, legal experts say it is likely to be influential.

"I think Judge Pham got it right," says Owen Jones, a law professor at Vanderbilt

University Law School in Nashville who attended the pretrial hearing. "He did a very thorough job, which will be useful for the courts that will have to confront similar issues in the future." Pham stopped well short of saying fMRI lie detection should never be admitted in court, however. In fact, says Jones, the report may give the companies developing this technology valuable guidance on the kinds of research that would improve their chances of getting it admitted in the future.

The government alleges that between 1999 and 2005, Semrau directed staff at two nursing homes he owned to submit fraudulent claims totaling about \$3 million. To convict Semrau, the government must show that he intended to cheat insurers. Defense attorneys assert that Semrau had no such intent, but was confused by unclear rules on reimbursement. They hoped to introduce fMRI scans performed by Cephus, a Tyngsboro, Massa-



Liar! fMRI scans may detect a neural signature of lying in the lab, but a federal judge says they're not ready for use in court.

chusetts, company, that purportedly show that Semrau is telling the truth.

But prosecutors requested a so-called Daubert hearing, named after a 1993 Supreme Court case that established guidelines used by most federal courts to weigh the admissibility of scientific evidence. These guidelines suggest that a given technology should meet four criteria: It should be subject to empirical testing, published in the peer-reviewed literature, have a known error rate, and be generally accepted in the

NEWSMAKER INTERVIEW: YOON DUK YONG

Crime Scene Investigation: The Sinking of the *Cheonan*

Simmering tensions on the Korean Peninsula are threatening to boil over following the South's accusation last month that a North Korean submarine torpedoed and sank one of its naval ships, killing 46 sailors. The North has angrily denied the charge. But "there is no other plausible explanation," contends Yoon Duk Yong, former president of the Korea Advanced Institute of Science and Technology (KAIST) in Daejeon, who co-led an investigation into the incident.

The 1200-ton *Cheonan* was ripped in two by an explosion near disputed waters in the Yellow Sea on 26 March. The two Koreas have had several naval clashes in the area since a 1953 armistice halted hostilities; technically they are still at war. But the latest incident may be the gravest. To determine what happened, the South Korean government assembled an international team of experts led by Yoon and a military investigator.

A native Korean, Yoon, 70, earned a bachelor's degree in physics from the Massachusetts

Institute of Technology and a Ph.D. in applied physics from Harvard University, both in Cambridge. He was on the faculty at Wayne State University in Detroit, Michigan, until joining KAIST in 1972. He served as KAIST president from 1995 to 1998 and now chairs the University Advisory Council at Pohang University of Science and Technology.

Yoon, a materials scientist, is not sure why he was tapped to head the investigation. Visual inspections, simulations, chemical analyses, and explosives testing all indicated that a torpedo exploded beneath the *Cheonan*, creating a massive bubble jet that burst through the hull. Then on 15 May, dredging crews found torpedo fragments at the site. "We have reached the clear conclusion that the *Cheonan* was sunk as the result of an external underwater explosion caused by a torpedo made in North Korea," Yoon reported at a press briefing in Seoul on 20 May.

According to news reports, South Korea plans to take the case to the U.N. Security

Council, though as *Science* went to press it was not yet clear what action it would request. North Korea has said that punitive actions could lead to war. With the stakes so high, Russia last week sent a delegation to South Korea to examine the evidence. Yoon says experts from China would also be welcome.

In an interview with *Science*, Yoon laid out the case against the North.

—DENNIS NORMILE

Q: How did the investigation proceed?

Y.D.Y.: We considered all possible causes, including grounding, collision, internal explosion, fatigue failure, and accidental attack by a friendly force. But these were ruled out one by one. The possibility of an explosion under the ship was supported by the deformation and fracture of the ship's hull.

In addition, there were traces of explosive chemicals and amorphous aluminum oxide formed from aluminum powder in the explosive. Aluminum powder is added to



Did Polynesians reach South America before Europeans did?

1344



U.K. libel laws and science

1348

scientific community. The Semrau case provided the first Daubert hearing for fMRI lie detection. “To me it’s very exciting,” says Martha Farah, a cognitive neuroscientist at the University of Pennsylvania who flew to Memphis to observe the hearing. “I mean, talk about the interface between neuroscience and society, this is it.”

Farah, like many neuroscientists, is deeply skeptical about using fMRI lie detection in legal cases, and she says she went into the hearing thinking there was no chance the judge would allow it as evidence. But after a day and a half of testimony, she wasn’t so sure. Cephus CEO Steven Laken “spoke in very strong and confident terms” about published research finding that certain regions of the brain become more active when a person lies, Farah said. (For her play-by-play of the hearing, see *ScienceInsider* <http://bit.ly/brain-scans>.) Laken also described the procedure Semrau underwent. In short, Cephus scanned Semrau’s brain as he answered questions about details of his case and compared those scans with others taken as he answered questions unrelated to the case, after being

instructed to intentionally lie on certain questions and respond truthfully to others.

Judge Pham concluded that fMRI detection appears to meet the first two Daubert criteria—that is, it is testable and has been described in several peer-reviewed articles—but falls short on the criteria regarding error rates and acceptance by the scientific community.

At the Daubert hearing, Peter Imrey, a biostatistician at the Cleveland Clinic Foundation in Ohio, provided a detailed critique of the impressively low error rates cited in Cephus’s promotional materials and in papers published by Laken and colleagues. Pham took note of this and also expressed concern that the available error rates come from carefully controlled laboratory experiments. “There are no known error rates for fMRI-based lie detection outside the laboratory setting, i.e. in the ‘real-world,’” he wrote.

As for acceptance of the method by the scientific community of fMRI-based lie-detection technology, Pham noted that even a 2008 paper in *Directions in Psychiatry* co-authored by two members of Cephus’s sci-

entific advisory board had stated that the method wasn’t ready for prime time: “The authors conclude by stating ‘Functional MRI is currently not ready to be used in real-world lie detection,’” Pham wrote.

Pham’s report emphasizes the disconnect between the published work on fMRI lie detection, which generally involves scans performed on college undergraduates instructed to lie about a mock crime, and Semrau’s case. Semrau was 63 at the time of his first scan, his alleged crimes happened several years earlier, and he faces serious consequences if convicted. More naturalistic experiments that try to account for such factors might help proponents of fMRI lie detection make a stronger case, say Jones and Farah.

Laken declined to comment specifically on Judge Pham’s ruling or what it means for Cephus’s plans to develop its technology. “This is just one ruling,” he says, noting that it doesn’t necessarily apply to other jurisdictions. In the week since the court decision was released, Laken claims Cephus has received an uptick in queries about its services.

—GREG MILLER

torpedo or mine explosives to increase the explosive and bubble effects. An acoustic signal detected at the moment of the incident indicated an explosion equivalent to about 260 kilograms of TNT. And a computer simulation showed an explosive effect of about 200 to 300 kilograms of TNT detonated at a point below the ship.

But the smoking gun was the torpedo [debris], consisting of two propellers, a motor, and steering wings. They were found close to the explosion point along with debris from the sunken ship. Furthermore, the oxidation of the steel pieces [of the torpedo] was roughly the same as that found on the fractured bow of the ship, and the same amorphous aluminum oxide produced by the explosive was found on the propellers and motor. The estimated explosive strength of the torpedo agreed with that expected to have caused the damage to the ship. The recovered parts agree exactly in size and shape with detailed drawings prepared by North Korea to export these torpedoes. Additional evidence was the writing of “No. 1” in Hangul, the Korean alphabet, on a metal plate near the screws. We concluded that these propulsion parts came from the torpedo that was

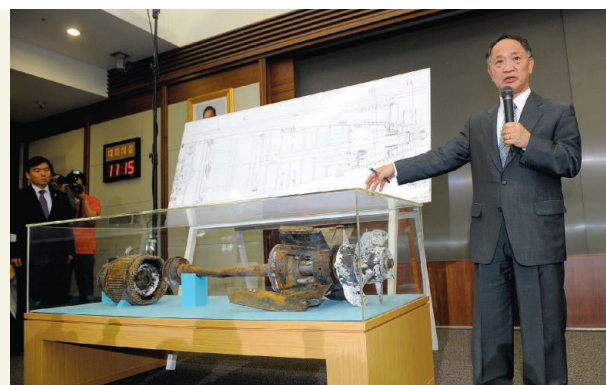
made by North Korea and used to attack the *Cheonan*.

Q: Without that smoking gun, how strong would your evidence have been?

Y.D.Y.: I think our conclusions would have been less definitive and convincing. We would have had to draw our conclusion on the basis of circumstantial and indirect evidence, which was still quite substantial.

Q: What did international experts bring to the investigation?

Y.D.Y.: They were invited to enhance the objectivity and credibility. Korea had never had a naval incident like this before. But these countries [Australia, Sweden, the United Kingdom, and the United States] had similar incidents dating back to the Second World War. Some members of the international team had investigated recent incidents involving their ships and those of other countries. Some are also knowledgeable of tests in which old ships were sunk by



Smoking gun. Materials scientist Yoon Duk Yong gestures toward the remains of a torpedo, apparently made in North Korea, that was found in the area where the *Cheonan* went down.

torpedoes. They contributed a great deal to setting the direction of the investigation for the *Cheonan* incident and reaching a unanimous conclusion.

Q: Have outside experts raised any questions about your methods or conclusions?

Y.D.Y.: There were no really serious objections about the methods. I think the physical evidence is so clear and definitive that any experts will agree with our conclusion.

CREDIT: JUNG YEON-JE, POOL/AP PHOTO

U.S.-INDONESIA AGREEMENT

U.S. Effort to Reach Out to Muslim-Majority Nations Begins to Bear Fruit

JAKARTA—Crisscrossed with fissures and vents, the sea floor off Java Island's northern coast is a patchwork of unusual ecosystems. In its first international mission, the U.S. research vessel *Okeanos Explorer* early this summer will team up with an Indonesian vessel, the *Baruna Jaya IV*, to probe the ecological hotbed.

The expedition ushers in a new era in science cooperation between Indonesia and the United States. The two countries have just inked their first S&T agreement, which is now awaiting ratification by Indonesia President Susilo Bambang Yudhoyono. And two high-profile initiatives are in the works. In the coming weeks, the United States is expected to unveil an extensive education package, including university partnerships and dedicated funds for S&T collaboration; funding for the package could top \$150 million. It will also tap Indonesia to host a regional center for climate change, one of the centers of excellence for the Muslim world that U.S. President Barack Obama promised to establish in a landmark speech in Cairo last year. (Both initiatives were to be announced this month during Obama's planned visit to Indonesia, which was postponed.)

Other signs of a closer relationship include an annual Frontiers of Science meeting that the Indonesian Academy of Sciences (AIPI) and the U.S. National Academies intend to launch next year to spark collaborations between top young scientists. And an Indonesian-U.S. team is now drilling ice cores from a tropical glacier (*Science*, 28 May, p. 1084). "This whole spectrum of activities will strengthen ties between our two countries," says Jason Rao, a senior policy analyst at the White House's Office of Science and Technology Policy.

Yudhoyono spoke of Indonesia's own efforts to bolster science in a January speech to AIPI, urging researchers to take risks and "be much more open-minded and more pro-

gressive" than in the past. He returned to the theme in a meeting in Jakarta last month with Bruce Alberts, *Science's* editor-in-chief and one of three science envoys appointed by Obama to explore collaborations with Muslim-majority countries. Discussions are also under way on creating a merit-based agency similar to the U.S. National Science Foundation (NSF). After years of stagnation, researchers here sense "the beginning of a renaissance," says medicinal chemist Umar Anggara Jenie, chair of the Indonesian Institute of Sciences (LIPI) in Jakarta.

The United States is playing a critical supporting role in this revival. In his speech, Yudhoyono cited cooperation in technology and education as key elements of a "new strategic partnership" between the two countries. Another impetus is Indonesia's problem with homegrown terrorists. In Cairo, Obama vowed to support technological development in Muslim-majority countries. "We understand his intention to bridge Islamic civilization and the West. Science is the best way to do this," says AIPI President Sangkot Marzuki, director of the Eijkman Institute for Molecular Biology (EIMB).

Indonesia's first biosafety level-3 lab—capable of handling the most dangerous pathogens—opened here recently at EIMB. Eijkman's forensics lab, meanwhile, has aided the government's antiterrorism efforts. "Indonesia has made tremendous strides in science since I first began working here in 1987," says J. Kevin Baird, director of the Eijkman-Oxford Clinical Research Unit. "There are definite new signs of scientific vibrancy." Another bright spot is Universitas Pelita Harapan, which is building a nanotechnology research center. And in Bogor, LIPI has just opened an International Center for Interdisciplinary and Advanced Research. The center, which aims to hire several hundred young research staff members, will focus on hot issues such as climate change, food security,



Deep relationship. The U.S.'s *Okeanos Explorer* (inset) and an Indonesian ship are joining forces this summer to explore the diverse ecology of the Coral Triangle.

green technology, and crisis management.

To stimulate innovation, Yudhoyono is expected to announce a national oversight committee chaired by Zuhul (who uses one name), rector of Universitas Al-Azhar and a former science minister. Also, a program soon to be launched will give block grants to industry for R&D. The idea is to "shift the mindset" that innovation only occurs in research institutions, says Teguh Rahardjo of the State Ministry of Research and Technology. The private sector, Rahardjo says, will be expected to contribute a rising share of the nation's R&D spending.

Huge challenges remain. Since Yudhoyono took office in 2004, the government's R&D budget has doubled to roughly \$200 million. Still, that's a paltry 0.06% of GDP. "How can you do excellent science on such little money? If you want to be an innovative nation, you can't do it on 0.06%," says LIPI materials scientist Sunit Hendrana. That has meant scant resources for merit-based grants. "I heard this over and over from young scientists," says Alberts, who spent 11 days in Indonesia last month. "Our most important need is a foundation like the NSF," says Marzuki, who says AIPI may try to set one up with private sources and matching U.S. funds.

A more fundamental concern is science education. The government intends to triple university enrollment in natural sciences to 12% by 2014. But students are ill-prepared: Primary school science education is woeful, researchers contend. Plans are under way to bring Indonesian educators to a workshop on inquiry science for children next month at the National Science Resource Center in Washington, D.C., run by the National Academies and the Smithsonian Institution. "The hope is to initiate an impressive program of science education in one or two carefully selected Indonesian school districts," says Alberts.

If the pilot project succeeds, Alberts says, "this would support Indonesian hopes to become an enlightened model for other Muslim-majority nations and beyond." To Rao, that's exactly what the enhanced S&T collaboration with Indonesia is meant to achieve.

—RICHARD STONE

CREDITS (LEFT TO RIGHT): DAVID R. BURDICK; NOAA

U.S. SCIENCE POLICY

Obama's Nominee to Lead NSF Lauded For Science and Management Skills

Intense but relaxed. Driven but calm. Supremely confident, yet self-effacing enough to gain the trust of those who might disagree with him.

A portrait of President Barack Obama? Some may think so. But the description comes from colleagues asked about the president's latest scientific appointment: Subra Suresh, dean of engineering at the Massachusetts Institute of Technology (MIT), who was nominated last week to become the next director of the National Science Foundation (NSF).

The 54-year-old Suresh isn't well known in Washington, D.C., science policy circles. But his ability to maintain a cutting-edge lab even as he ascended the administrative ranks at MIT—hired as a professor in 1993, he became head of the department of materials science and engineering in 2000 and dean

in 2007—is thought to have put him at the top of the list of candidates compiled by John Holdren, the president's science adviser (*Science*, 19 March, p. 1438).

Suresh stood out even at elite institutions like MIT, where he finished his doctorate in mechanical engineering in 2 years, and at the University of California, Berkeley, where he did a postdoc, says Robert Ritchie, his advisor at MIT, who brought Suresh with him when he moved to UC Berkeley. "The thing about Subra is that, while he'll take on tasks that are quite complex, you know from the start that he'll succeed," says Ritchie. "And he'll make sure that his other duties don't suffer. I think he's the perfect candidate for the [NSF] job."

Born in India and educated at the Indian Institute of Technology, Madras, Suresh has moved into new research fields without miss-



Lucky 13? Subra Suresh would be NSF's 13th director.

ing a beat. It's allowed him to go from studying fatigue in large structures such as airplane wings to probing the characteristics of microelectronic materials to understanding the biomechanics of malaria-infected human red blood cells. "It's not interdisciplinary research," explains Ben Freund, an engineering professor at Brown University and a for-

SCIENCE FUNDING

Merger to Create New Voice for European Science

In an effort to simplify the alphabet soup of the continent's research landscape, some pan-European bodies have begun planning a merger. The aim is to create a single, and more powerful, voice for the national science funding agencies that control some €25 billion annually—by far the largest slice of public money available to the region's scientists (see chart). The heads of Europe's national research councils, who gather regularly under the soubriquet EUROHORCS, look set to join forces with the European Science Foundation (ESF), which funds research, backs research networks and conferences, and develops sci-

ence policy, mostly with money from EUROHORCS members. "We will try to combine the best things that characterized the two organizations in the past and have the courage to leave out others," says Dieter Imboden, current EUROHORCS president. The new body, which could begin work next year, has been given the tentative working title of the European Research Organization (ERO).

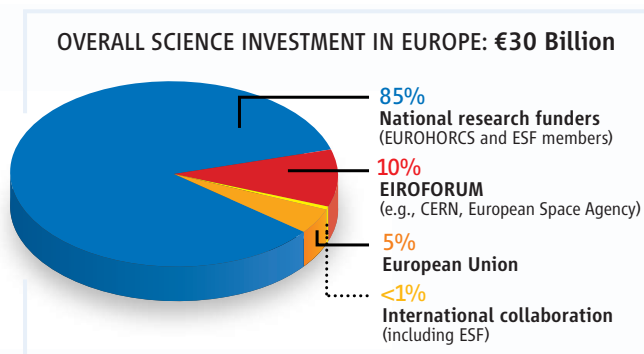
The move has been prompted in part by the growing influence of the European Union over the continent's research policies, in particular its efforts to harmonize national science career structures and funding systems so that researchers can move freely from country to country—a goal known as the European Research Area (ERA). "Europe has to be a little bit like the U.S. and get this very fluid movement of people and money," says Arnold Migus, former director general of France's research agency, CNRS.

The national funding bodies of EUROHORCS, which also includes ones

from non-E.U. states such as Norway and Turkey, came late to the ERA bandwagon. After the European Union published a discussion document on the ERA in 2007, EUROHORCS decided to support the process and wanted to influence it. EUROHORCS teamed up with ESF to draw on its policymaking expertise, and the two bodies produced their own road map of actions to achieve an ERA. These include allowing researchers to take funding with them if they move to an institution in another country, harmonizing peer review and career structures, promoting open access to results, and making it easier to create cross-border infrastructures.

The problem was that, although EUROHORCS members provide about 85% of Europe's public research funding, they were not getting their collective message across. "We realized that the voice from EUROHORCS and ESF was not very clear," says Pär Ommling, head of the Swedish Research Council and a former EUROHORCS president. From the outside, and particularly from the E.U. government's point of view, it wasn't clear why there are two organizations.

ESF was created in the 1970s by a few funders that wanted to collaborate on certain projects. Since then it has grown to include



Big slice. A new European organization formed by a merger will represent some €25 billion in science funding.

ScienceNOW

mer dean who hired Suresh for his first faculty position and who later co-authored a popular textbook with him on thin films. "It's moving from the interface of two disciplines to the interface of two other disciplines."

Suresh's career is studded with such collaborations. Ares Rosakis, chair of the division of engineering and applied sciences at the California Institute of Technology in Pasadena, savors a 25-year research partnership that began when Suresh was an assistant professor at Brown and includes a short-lived start-up company to provide diagnostic tools for the semiconductor industry. "That can put a strain on any relationship," says Rosakis about the company, Oraxion Diagnostics, he and Suresh launched in 2003 and sold 3 years later to a private company. "But he's such a delightful partner even that experience went smoothly."

Those interpersonal skills have also served him well as an academic administrator. Freund says Suresh's efforts to strengthen course offerings within MIT's materials science department, for example, "would have

caused a revolt among most faculty." But Suresh "got people to buy into the changes." And David Parks, a professor of mechanical engineering at MIT who has known Suresh for 30 years, says "I don't think I've ever seen him in anything other than good humor. His attitude is, 'Let's get this done.' He's very confident. And his presentations are so clear and well done that after hearing them, you just want to go along."

Ritchie and others are betting that Suresh will continue to do science even as he runs the \$7 billion NSF, which supports fundamental research across all scientific and engineering disciplines. "I think he'll find a way," says Ritchie. If so, he would be possibly the first NSF director to continue an active research program. Confirmation by the U.S. Senate would make Suresh the 13th director of the 60-year-old agency, the fifth engineer, and the first of Asian descent. Acting Deputy NSF Director Cora Marrett has succeeded Arden Bement, who returned to Purdue University last week to lead a new global science policy institute.

—JEFFREY MERVIS

79 organizations, including bodies that fund research and bodies that carry it out, as well as Western-style science academies, which are essentially learned societies. With its modest annual budget of about €50 million, ESF has never carved out a role as the preeminent European funding body. "It's been kept on a short leash by its members," says Helga Nowotny, president of the European Research Council, an E.U. funding agency. "It's had a hard time looking for a profile."

The members of EUROHORCS, on the other hand, control large budgets, but the organization has no headquarters or staff. "EUROHORCS has gained momentum. It's now a partner with [the E.U.], but it's a bunch of generals without soldiers," says Imboden, who also heads the National Research Council in Switzerland. "Faced with this situation, we could either found our own secretariat or use the ESF."

The current plan of taking the latter course will not be simple. EUROHORCS would like an office in Brussels and ESF is headquartered in Strasbourg with some 125 staff. In addition, ESF includes academies that are learned societies, but "EUROHORCS really wants an organization of funding agencies," says ESF President Ian Halliday. Working groups are hammering out the details of the merger for approval by EUROHORCS and



"European science is badly served." ESF's Ian Halliday.

ESF's assembly in the autumn.

One major change is that ESF's grant-giving function will cease. According to Halliday, the new body will be much stronger in policy advice, in foresight—setting the agenda for European science—and in joint programming. ERO will coordinate the efforts of different national agencies in key areas to avoid duplication. For example, in energy research, Halliday says, there's no strategic plan across the region: "European science is badly served."

After a meeting of ESF's governing council last month, the two bodies have a mandate to move the merger forward. "We are in a process. A year ago it looked like mission impossible, and the situation is not solved yet," says Imboden. "I'd be extremely surprised if it's not successful," says Omeling. "The momentum is there."

—DANIEL CLERY

From *Science's* Online Daily News Site

Learning the Mongoose Ways

The select club of brainy critters known for carrying on traditions—among them humans, primates, whales, and dolphins—has an unlikely new member: the banded mongoose.

Researchers have found that the furry African carnivore learns by imitation as well and carries what it learns well into adulthood.

Experts say the discovery offers the first direct observation of animals passing down traditions in the wild.

<http://bit.ly/mongoose-ways>



Crocodiles Know How to Ride the Tides

The giant crocodiles that ply Australia's rivers and bays are also savvy ocean voyagers: The 5-meter-long reptiles make good use of ocean and river currents to travel hundreds of kilometers, according to one of the most extensive monitoring studies of the animals ever attempted. And like experienced sailors, they prefer to navigate when the tides are in their favor. <http://bit.ly/crocs-know>

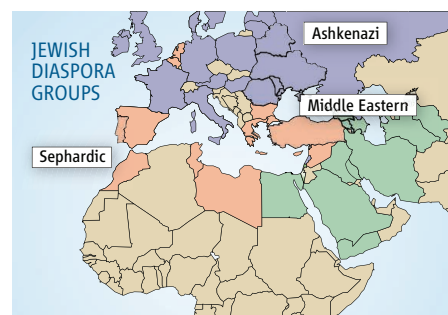
Elusive Particle Caught in Strange Appearing Act

Like a magician pulling a rabbit from a hat, an international team of physicists has made one kind of neutrino appear from the beam of another type of neutrino. Their odd observation clinches the case that these elusive particles can transform into another in a process called neutrino oscillations, an idea that had gained wide acceptance in recent years but had not been directly observed. <http://bit.ly/elusive-particle>

Tracking Disease Through Mosquito Slobber

This is nothing to spit at: Scientists say they may be able to track deadly mosquito-borne diseases by studying the saliva the insects leave behind when they feed on sugary bait. <http://bit.ly/mosquito-slobber>

Read the full postings, comments, and more at news.sciencemag.org/sciencenow.



Far-flung family. Ethiopian Jews (left) are genetically distant from others but observe Jewish customs.

HUMAN GENETICS

Who Are the Jews? Genetic Studies Spark Identity Debate

The world's 13 million Jews are strongly linked by religion and culture. But do they share a common genetic heritage? Two new studies conclude that most members of the far-flung Jewish Diaspora can trace their roots to ancestors who lived in the Middle East more than 2000 years ago. The new research, based on recent advances in genome technology, apparently refutes controversial claims that most of today's Jews descend from more recent converts. And it finds that Jews in Ethiopia and India who also claim origins in ancient Israel are more distantly related to other Jewish groups. Yet some researchers argue that although science can track Jewish ancestry, it has little to say about who is a Jew today.

The studies "clearly show a genetic common ancestry" of most Jewish populations, says Sarah Tishkoff, a geneticist at the University of Pennsylvania, thus indicating a distinct Jewish people through history. Indeed, says Harry Ostrer, a geneticist at New York University Medical School and leader of one of the teams, the genomewide scans used in the studies can detect Jewish ancestry in anonymous DNA samples. But Doron Behar, a geneticist at the Rambam Health Care Campus in Haifa, Israel, and lead author of the second report, argues that genes do not necessarily make the Jew. There is no "metaphysical" difference between someone born Jewish and a convert to Judaism, Behar says.

The two studies—one led by Behar and published online this week in *Nature*, and Ostrer's, published last week in *The American Journal of Human Genetics*—speak to a current debate about Jewish origins, including that of the Ashkenazi Jews of Europe, who

make up 90% of American Jews and nearly 50% of Israeli Jews. Tel Aviv University historian Shlomo Sand's 2008 book *The Invention of the Jewish People* argued that few modern Jews can trace their heritage to ancient Israel. He in part resurrects a thesis, made famous by writer Arthur Koestler in the 1970s, that Ashkenazi Jews are actually descended from a Turkic people in Central Asia whose rulers converted to Judaism in the 8th century C.E.

The new studies contradict that conclusion. Both teams used DNA microarrays to examine variation within Jewish groups worldwide and between those groups and non-Jewish populations. Microarrays allow comparisons of thousands of genetic differences, from single nucleotide pairs to longer stretches of DNA, between different individuals (*Science*, 21 December 2007, p. 1842). This genomewide approach is much more powerful than previous analyses of Y chromosomes and mitochondrial DNA, which were often inconclusive.

The Ostrer team analyzed nuclear DNA from 237 Jews representing the three main Diaspora groups: Ashkenazi Jews; Sephardic Jews from Spain and Portugal; and Middle Eastern, or Oriental, Jews. Their DNA was compared with that of about 2800 presumably non-Jewish people from around the world. The Behar study employed smaller sample sizes—121 Jews and 1166 non-Jews—but from more population groups and also analyzed 8000 non-Jewish Y chromosomes and 14,000 mtDNA genomes.

The studies came up with very similar results: Jews from the three Diaspora groups were closer to each other genetically than to non-Jews from the same geographic region.

Indeed, the Ostrer study found that Ashkenazi Jews were as closely related to each other as fourth or fifth cousins, even though their genetic profiles indicated between 30% and 60% admixture with non-Jewish Europeans. Jewish groups also clustered tightly with non-Jewish Middle Eastern populations such as the Druze and the Cypriots, strongly suggesting an origin in that geographic region. "I would hope that these observations would put the idea that Jewishness is just a cultural construct to rest," Ostrer says.

The Behar study also included three Jewish groups whose ancestry has been uncertain: the Beta Israel Jews of Ethiopia, the Cochin Jews of southern India, and the Bene Israel Jews of northern India. It found that all three groups genetically clustered with non-Jewish Ethiopians and Indians rather than with the Diaspora groups. However, analysis of the Y chromosomes of the Bene Israel Jews showed paternal links to the Middle East, suggesting that they might share ancient roots with Jews from that region. So it's possible, Behar and other researchers say, that the Ethiopian and Indian Jewish groups were founded by Jews from other regions who then intermarried and/or converted many local non-Jews to Judaism, thus expanding their numbers but diluting their Jewish genetic signatures.

Overall, "these results confirm the common wisdom that Jews have always held," that they stem from a common Middle Eastern origin and heritage, says historian Anita Shapira, also from Tel Aviv University. "It is nice to get support from modern genetics, which refutes [Sand's] assertions," she says.

Sand counters that the whole concept of identifying Jews genetically is fallacious. "No study ... has succeeded in identifying a genetic marker specific to Jews," he insists. He adds, "It is a bitter irony to see the descendants of Holocaust survivors set out to find a biological Jewish identity. Hitler would certainly have been very pleased." But Ostrer says that Sand has not kept up with advances in genetic research: "We can tell who the Jews are genetically." —MICHAEL BALTER

CREDITS (LEFT TO RIGHT): THE ISRAELI ASSOCIATION ON BEHALF OF JEWISH ETHIOPIA/WIKIMEDIA COMMONS; ADAPTED FROM H. OSTRER, NATURE REVIEWS GENETICS, 2 (NOVEMBER 2001)

Downloaded from www.sciencemag.org on June 10, 2010

PSYCHOLOGY

The Prickly Side of Oxytocin

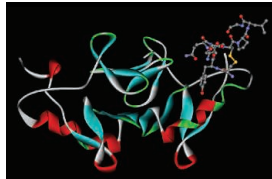
Oxytocin has a touchy-feely reputation, thanks to research showing that it promotes social bonding in a wide range of animals, including humans. Sold on the Internet in a formulation called “Liquid Trust,” the peptide hormone is marketed as a romance enhancer and sure ticket to business success. Australian therapists are trying it alongside counseling for couples with ailing marriages. And police and military forces reportedly are interested in its potential to elicit cooperation from crime suspects or enemy agents.

But a study published on page 1408 hints that the hormone has a prickly side as well. In experiments with groups of people playing an economic game, those who received a dose of oxytocin behaved more altruistically toward members of their own group. Yet they also displayed more “defensive aggression” toward outsiders, preemptively punishing members of a competing group when their own group was in danger of suffering a heavy financial loss. “The important message here is that oxytocin is not just promoting generosity and benevolence and trust,” says lead author Carsten De Dreu, a psychologist at the University of Amsterdam in the Netherlands. “It’s a double-edged sword.”

De Dreu and colleagues recruited dozens of undergraduate men to play a version of the prisoner’s dilemma game that’s often used in behavioral economics research. Half an hour before the game, all participants inhaled three puffs from a bottle of nasal spray containing either oxytocin or a placebo. (Neither they nor the researchers guiding them through the experiment knew which.) When the game began, participants were assigned to groups of three and told that another group of three would be playing at the same time. Each individual received €10 (about \$13) and had to pick from several options for distributing his money. Compared with men who got the placebo, those who’d taken oxytocin behaved more altruistically, keeping less money for themselves and donating about twice as much to a

pool that benefited all members of their own group.

However, a more complicated version of the game revealed oxytocin’s darker side. De Dreu’s team tested 75 men in a game in which groups of three sometimes faced the possibility of a substantial financial loss to their group if the other group decided not to cooperate with their group by giving them money. This threat didn’t faze men who’d taken the placebo, but the oxytocin-sniffers tended to get defensive, selecting an option that took money from the outsiders and minimized their group’s potential losses instead of playing nice and hoping the others did the same. Oxytocin had this effect only when



Not so cuddly? Oxytocin (below) may help mothers bond with their young—and defend them in times of danger.

men faced a significant financial threat to their group; otherwise, those who took the hormone behaved no differently toward outsiders.

The findings hint at a biological basis for thinking that altruism and aggression are more closely related than is usually acknowledged, says Holly Arrow of the University of Oregon, Eugene, who studies group dynamics and the psychology of war. “It’s morally simple to say that altruism is good and aggression is bad,” Arrow says. But in the context of war, for example, harming others to protect one’s own group can be considered heroic. “Oxytocin is perhaps an important pathway that bonds men together and makes them ready to defend the group,” Arrow would like to see the work repeated in

women to see if they respond the same way.

Oxytocin regulates mother-infant bonding in sheep and rodents, pair bonding in voles, and flock size in zebra finches (*Science*, 14 August 2009, p. 862), and Larry Young, a neurobiologist at Emory University in Atlanta, says De Dreu’s study suggests that the molecule plays a similarly important role in maintaining the more complex social groups formed by humans. “It probably does have a big role in our everyday relationships and how we fit into groups and contribute to the success of our own group,” Young concludes.

—GREG MILLER

Science Insider

From the *Science* Policy Blog



Italian researchers are worried that their successful protests against **budget cuts** proposed by Prime Minister Silvio Berlusconi may be only a temporary victory. <http://bit.ly/italyfunts>

A scientific meeting last week in Washington, D.C., is the latest sign that the National Football League has reversed its longtime resistance to acknowledging a link between the sport and **brain injuries**. <http://bit.ly/concussions>

As part of our team coverage of the **Gulf of Mexico oil spill** (see <http://news.sciencemag.org/oilspill>), we reported:

- A story on how scientists are stunned by a rush analysis of a **hand-drawn engineering plan**;
- How scientists with the National Oceanic and Atmospheric Administration’s toxics lab in Seattle, Washington, are **looking for oil in seafood samples and also shifting their strategy** to begin sampling in oiled areas as well as clean areas;
- The recent discovery of pancake batfish 400 meters below the surface illustrates how little scientists know about species abundance on the sea floor—and just what might be at **risk from the oil**.
- NOAA Administrator Jane Lubchenco is calling for a halt to **missions to find oil plumes** in the gulf until coordinated research can be planned in a “careful, thoughtful, and quantifiable manner”;
- One oceanographer who uses undersea robots to study jellyfish and other sea life says that the **moratorium on deep-sea drilling** in the gulf has cut off his flow of data. He collects images from undersea cameras connected to the robots that work below the drill platforms;
- And new modeling supports previous predictions that the oil will likely move around Florida and **enter the Gulf Stream**.

See the full postings and more at news.sciencemag.org/scienceinsider.

Beyond *Kon-Tiki*: Did Polynesians Sail to South America?

After decades of taboo and controversy, Pacific Rim archaeologists are finding new evidence that Polynesians reached South America before Europeans did, voyaging across the world's largest ocean around 1200 C.E.



IT WAS LATE IN THE DAY IN DECEMBER 2007 when the curator pulled out a half-dozen ancient human skulls from dusty drawers in the museum in Concepción, Chile. The skulls, of varying ages, had been found during the past century by locals as well as scientists on the windswept island of Mocha, 30 kilometers off the southern Chilean coast. “I nearly dropped to the floor,” recalls Lisa Matisoo-Smith, a biological anthropologist at the University of Otago in New Zealand. She immediately noticed that some crania had characteristics hinting at a Polynesian origin, such as a pentagonal shape when viewed from behind.

Online sciencemag.org

S Podcast
interview
with author
Andrew Lawler.

Mocha is 3700 kilometers east of Easter Island, the closest known prehistoric Polynesian settlement. Matisoo-Smith was in Chile on a hunt for rat bones that might show Polynesian contact with South America; she hadn’t imagined stumbling on human remains that might bolster that case. She and Chilean colleague José-Miguel Ramírez-Aliaga of the Universidad de Valparaíso hope to win agreement from local peoples and the Chilean government for an excavation on Mocha to seek signs of Polynesian settlement, artifacts, and human and animal remains.

Their effort is one part of an ambitious

drive to settle a long-standing controversy among archaeologists and anthropologists. Considered the realm of crackpot theorists until recently, the idea of prehistoric contact between Polynesians and South Americans has gone mainstream. A new generation of researchers is using DNA analysis of varied organisms such as humans, chickens, and sweet potatoes to add compelling data to a case previously based on more nebulous linguistic and artifact similarities. Given current views of Polynesian expansion (see sidebar, p. 1346), many researchers now think it likely that Polynesians reached South America by about 1200 C.E., after the settlement of Easter Island, and several centuries before Europeans arrived around 1500 C.E.

“This is a watershed moment,” archaeologist James Bayman of the University of Hawaii, Manoa, told participants at a session on the topic at the meeting of the Society for American Archaeologists (SAA) in St. Louis, Missouri, in April. “New methods no longer give us an excuse to ignore the issue.” Some skeptics point out that there is still no incontrovertible evidence that Polynesians went to South America and then returned to Pacific islands, and contact with North America remains questionable (see sidebar, p. 1347). Bayman admits that the research “is still a work in progress.” But he and many colleagues agree that resistance to the idea of prehistoric contact is starting

to crumble, giving archaeologists a chance to rethink the way technology and innovations spread in prehistory.

Heyerdahl’s ghost

The idea that Polynesians and South Americans were in touch more than 500 years ago is as old as archaeology itself. In 1837, a French writer noted that plank boats used by locals on the west coast of Chile were remarkably similar to those found in Tahiti. Two years later, a British sea captain pointed out that the Patagonian and Polynesian words for canoes—*kialu* and *kialoa*, respectively—were nearly the same, notes Kathryn Klar, a linguist at the University of California (UC), Berkeley. An archaeologist suggested in the 1930s that the sweet potato diffused from its home in the Andes to the Pacific before Columbus. And researchers at a 1968 conference concluded that the pre-Columbian presence in Polynesia of indigenous South American plants like the sweet potato were likely signs of prehistoric contact.

There is no question that the Polynesians were the great premodern seafarers, spreading east from Asia and Melanesia in outrigger canoes and arriving on the shores of Fiji by 1000 B.C.E. (*Science*, 2 March 2001, p. 1735). By 1200 C.E., using sails and sophisticated navigation techniques, they had peopled most South Pacific islands, including Hawaii and Easter Island on the

Seaworthy. The reach of Polynesians, shown here in this 18th century engraving, may have extended as far east as South America.

eastern edge of the Pacific. But did they go all the way to South America?

Scientists largely blame Thor Heyerdahl, the famous Norwegian writer and adventurer, for souring academia on that idea. In 1947, Heyerdahl and a crew of five sailed nearly 7000 kilometers from Peru to French Polynesia on a balsa-wood raft dubbed the *Kon-Tiki* to prove that ancient Americans could have sailed east to colonize the Pacific. The trip and resulting book made him a worldwide celebrity. Heyerdahl thought that the original seafarers came from the Middle East via South America, where they bestowed civilization on dark-skinned peoples before setting off across the world's largest ocean.

Such racist assumptions and a lack of scientific rigor horrified many anthropologists and tarred researchers who wanted to examine such prehistoric long-distance connections more thoroughly. "People asked me if I wanted to ruin my career and be considered a fool," recalls Terry Jones, now a professor at California Polytechnic State University in San Luis Obispo. The topic is still controversial, though no longer taboo. "We have to tiptoe back and reexamine just what the connections are," says archaeologist Terry Hunt of the University of Hawaii, Manoa.

The most compelling evidence, scholars say, centers on the humble sweet potato. That tuber is widely recognized to have been domesticated only once, about 6000 B.C.E.



Stone age. Ram3rez-Aliaga (right) notes that M3ori hand clubs (top row) resemble Mapuche versions from Chile.

in the uplands of Peru, where it became an important staple, according to Andrew Clarke, a molecular biologist at the University of Otago. Unlike the coconut or the gourd, which can naturally float from island to island, "the sweet potato needs people" to spread, he says. Some scholars in the past argued that the sweet potato was exported to Southeast Asia by the Spanish and Portuguese in post-Columbian times, then spread east across the Pacific. But the tuber, still a mainstay of the Polynesian diet, has shown up frequently in much earlier Pacific sites. Patrick Kirch of UC Berkeley, for example, dated a carbonized sample from the Cook Islands northeast of New Zealand at about 1000 C.E.

Such ancient samples have been known for decades but received little attention. Now

genetic tools are clinching the argument that sweet potatoes were brought to Pacific islands before the advent of Europeans. "It is easy to spot changes in the sweet potato genome over time," explains Clarke, who presented some of his data at the SAA meeting. That's because his team of researchers from New Zealand and Japan is using a high-resolution molecular marker technique to illuminate the large amount of genetic variation found in the plant. They examined 300 samples collected from around Oceania, South America, and Southeast Asia and found that the varieties common in Polynesia differ from those brought by Europeans to Southeast Asia; the Polynesian varieties are more directly related to the South American ones, Clarke says, strongly suggesting prehistoric contact.



Chicken feed

Just how a highland Andes crop appeared in Polynesia in pre-Columbian times remains controversial. "The sweet potato shows movement from South America to Polynesia, but not how it happened or who was involved," says Atholl Anderson, an archaeologist at the Australian National University in Canberra. Given the crop's highland origin and the prevailing winds, he argues that it is more likely that Amerindians dispersed the tuber to Pacific islands, moving west as Heyerdahl had proposed.

But most researchers see few signs of Amerindian excursions into the Pacific. Cultural anthropologist Richard Scaglion of the University of Pittsburgh in Pennsylvania instead argues that Polynesians may have arrived at the southern coast of South America and sailed north using the prevailing current to the Ecu-

dorian port of Guayaquil, the only sheltered port in South America north of the rocky southern coast of Chile. The Canara people once lived from this area of the coast into the Andes highlands, making the sweet potato accessible to coastal visitors. And here the current veers sharply west. Computer simulations show that “the most successful return [westward] from the coast would be from Guayaquil to Polynesia,” Scaglione adds, making this area “a possible locus of trans-Pacific contact.”

Climate change may also have played a role in prehistoric contact. Heyerdahl argued that South Americans went west to colonize Polynesia in part because the winds at mid-latitudes blow from east to west. But paleoclimatic data from a 2002 study suggest that prolonged El Niño events reversed those winds around 1000 C.E., providing a window for Polynesians to voyage more easily to the South American coast. “The change in winds enabled them to go in new directions,” says Hunt.

Archaeological evidence for Polynesians in the area around Guayaquil—or anywhere else in South America—remains elusive. But researchers note intriguing similarities between objects found along the coast and those of a similar vintage common on distant Pacific islands. Some, such as stone fish weirs or fish hooks, could have evolved similar shapes independently, says Ramírez-Aliaga. But other artifacts from the Mapuche culture, centered on south-central Chile



Land ho. Mocha Island is drawing the interest of archaeologists seeking Polynesian remains in South America.

and including the area near Mocha Island, are strongly reminiscent of Polynesia. Polished stone hand clubs, wide at one end with a rounded short handle, look like wooden clubs used by the Māori people of New Zealand, who are of Polynesian descent. Stone clubs from the Chatham Islands off New Zealand’s eastern coast also look similar to another Mapuche type; both resemble stylized birds. And a Mapuche stone ax called a *toki*—the same word is used in Polynesia—resembles adzes used in Polynesia.

Some cultural traditions are also surprisingly similar. Both the Mapuche and the

Polynesians celebrated the New Year with the rising of the Pleiades after the winter solstice and used a magic *toki* to cut trees. Both play a game similar to field hockey, *pai pai* in the Austral Islands and *palin* in Mapuche.

Despite the sweet potato’s strong evidence, so far there is no evidence for other common Polynesian animals, like the dog and rat, in South America. But there are signs of pre-Columbian chickens. Domesticated in Southeast Asia, the chicken spread to Europe and was thought to have arrived in the Americas with Columbus. Then in 2007, a team

led by Alice Storey of the University of New England in Armidale, Australia, dated three chicken bones from El Arenal in Chile, about 100 kilometers from Mocha Island, to between 1300 C.E. and 1450 C.E. The claim grabbed headlines around the world.

But geneticist Jaime Gongora of the University of Sydney in Australia isn’t convinced by those few bones, saying that “a large number of specimens” need to be found and dated in independent laboratories to ensure reliable radiocarbon results. Storey says that she has new data that will soon be ready for publication.

Changing Time in the South Pacific

“When this comes to be prov’d we Shall be no longer at a loss to know how the Islands lying in those seas came to be people’d ... so we may trace them from Island to Island quite to the East Indies.”

—James Cook, 1769

Little did British explorer James Cook know that the question of when and how the Polynesians peopled the Pacific would still be a hot topic more than 2 centuries later. Most archaeologists now agree that Polynesians can trace their origins to the Lapita Culture, a fusion of Melanesian and Austronesian peoples that had spread as far east as Fiji by 1000 B.C.E. But archaeologists are deeply at odds over when Polynesians fanned out across the vast northern, central, and eastern Pacific Ocean.

One group favors a slower and earlier settlement that puts them on Hawaii as early as 500 C.E., while another argues that it wasn’t until 1200 C.E. that most of the far-flung islands were colonized. For the moment, those who argue for a swift and late settlement seem to have the upper hand. “But this is going to be a hard issue to resolve,” says archaeologist Terry Jones of the California Polytechnic State University in San Luis Obispo.

Many of the early radiocarbon dating samples collected during the past 6 decades are suspect, says Terry Hunt of the University of Hawaii,

Manoa, and the reliable dates are all later. Hunt argues that a host of early dates depend on carbonized remains found in sediment cores that could be the result of natural rather than humanmade fires or were done on driftwood, which could be centuries earlier than the people who used it. After carefully examining more than 1000 radiocarbon dates, Hunt believes that Polynesians may have reached East Polynesia sometime after 1000 C.E., in “an explosive dispersal” that extended to Easter Island, Hawaii, and possibly the coast of South America (see main text, p. 1344). It was “the most rapid phase of human expansion in world prehistory,” he says.

Not everyone is willing to let go of the longer chronology. Marshall Weisler, an archaeologist at the University of Queensland in Brisbane, Australia, complains that “there are pre-1200 C.E. dates for several island groups that are routinely ignored or dismissed by the short-chronology folks.” Many scientists, however, seem to be drifting in Hunt’s direction. “Increasingly, there is evidence for a rapid radiation around 1000 C.E.,” says Patrick Kirch of the University of California, Berkeley.

The key may lie in the Society Islands, which likely were an important point for Polynesian dispersal north to Hawaii and east as far as Easter Island and possibly South America. Two domesticated coconuts from the islands date to 600 C.E., but the earliest settlement appears to be centuries later. That ambiguity is a sign that archaeologists have more work to do before Cook’s musing can be definitively answered.

—A.L.



Contact clues. This Araucana chicken and these New Zealand sweet potatoes (*below*) may be legacies of Polynesian contact with South America.



Fiction to fact?

Sweet potato and chicken data will likely not be enough to convince skeptics. “We should be pursuing other lines of evidence,” such as Polynesian settlements in South America and ancient DNA, says Hawaii’s Hunt. “Human evidence would be the key.”

The skulls that Matisoo-Smith examined in Concepción provide intriguing hints of Polynesian ancestry. But DNA data would be more conclusive, and to date, few studies have sought Polynesian genetic markers in modern South Americans. Analyses of mitochondrial DNA in indigenous South Americans so far show no sign of a Polynesian incursion, says Matisoo-Smith. If contacts were minimal, those markers may be hard to find in living people. And because people of Polynesian ancestry settled in South America after Europeans arrived, researchers will need to look at DNA from ancient South Americans. Given that most prehistoric voyagers were likely male, scientists say they need uncontaminated nuclear DNA from an archaeological site in order to pinpoint the Y chromosomes of a Polynesian.

Mocha so far is the best candidate for such evidence and for signs of a Polynesian settlement on the continent. “A wider excavation will allow us to look for the settlement pattern, more burials, and of course more artifacts” from any Polynesian colony, says Hunt. But first the team is working with Mapuche leaders, the New Zealand embassy, and the Chilean government to take into account sensitivities among indigenous peoples in the Americas about archaeological digs.

Northern Exposure in Doubt

In the 1930s, famed anthropologist Alfred Kroeber noted that the Chumash Indians of Southern California made sophisticated sewn-plank boats remarkably like those constructed in Hawaii more than 4000 kilometers to the west. He suggested prehistoric Polynesian contact as the source of the Chumash technique. Now an archaeologist and a linguist are seeking to prove that old theory. But while researchers are making strides in demonstrating a connection between South America and Polynesia (see main text, p. 1344), the idea that California Indians learned from Hawaiians faces an uphill struggle. Questions about timing make many archaeologists skeptical.

The sewn-plank boats built by the Chumash reached more than 8 meters in length and could carry a dozen people, shuttling between the California coast and the Channel Islands, 30 kilometers offshore. Although smaller than the oceangoing boats built by Polynesians, they surpassed the technology of all other natives along the North American coast prior to the arrival of Europeans. The first hard evidence for the Chumash vessels appears circa 700 C.E., around the time that some say Polynesians were settling the central and eastern Pacific.

Terry Jones, an archaeologist at California Polytechnic State University in San Luis Obispo, believes that the appearance of the boats as well as Polynesian-style fishhooks in the same era provides convincing evidence for contact. And some linguistic evidence backs up that view. His collaborator Kathryn Klar, a linguist at the University of California (UC), Berkeley, has cataloged several Chumash words, including *tomolo*—the word for canoe—that appear to be closely related to Polynesian languages. Woodworking terms are also similar.

But others aren’t yet convinced. For example, it’s possible those words are not ancient and arrived with Europeans who had traveled in the Pacific in the 18th and 19th centuries. “Corned beef in a can has an indigenous Polynesian name,” archaeologist Terry Hunt of the University of Hawaii, Manoa, points out. And an increasingly influential group argues that Polynesians didn’t arrive in the eastern Pacific until 1000 C.E. (see sidebar, p. 1346), which would make the earlier Chumash innovation necessarily indigenous. “If they don’t arrive by 700 A.D., then it doesn’t fit,” says Patrick Kirch of UC Berkeley. Jeanne Arnold of UC Los Angeles adds that there is “zero archaeology for a Polynesian incursion” in North America. Finding Polynesian artifacts “would be wonderful,” Arnold says. “But I want to see the evidence.”



Tech transfer? Researchers debate whether this ancient sewn-plank canoe design from California is the result of Polynesian influence.

—A.L.

The work by young researchers like Clarke, Storey, and Matisoo-Smith is a sign that the taboo put in place a half-century ago in the wake of *Kon-Tiki* has lost its power. Clarke, for example, is eager to push on to study the bottle gourd, the tomato, soapberry, the coconut, and other plants that may have moved across the Pacific before European ships arrived. But that work still holds little interest to most scholars who focus on the Americas. “There’s been a glass wall separating the two regions,” says Hawaii’s Bayman.

That is changing, says Ramírez-Aliaga, who experiences less resistance among his South American colleagues to the idea of contact than in the past. Trade clearly was a two-way street, “so this takes nothing

away from native American groups,” says Matisoo-Smith.

Skeptics like Anderson and Gongora insist that much more data is necessary before they will accept the idea of seafaring Polynesians trading with ancient South Americans, and that scenario remains for now absent from world-history textbooks. Bayman cautions that overthrowing entrenched views will require additional lines of decisive evidence. But many Pacific Rim scientists say it is only a matter of time before a once-heretical notion becomes accepted wisdom. “When you put all of it together, I don’t see how you can interpret this any other way,” says Jones. “This is moving from compelling to accepted truth.” —ANDREW LAWLER



SCIENTIFIC PUBLISHING

A Chilling Effect?

There are calls for reforms to British libel laws after researchers were sued in the United Kingdom for discussing or writing about controversial matters

When Peter Wilmschurst boarded his flight to Washington, D.C., in October 2007, the British cardiologist could have had no idea that the lecture he was about to give would provoke a libel suit—or that 3 years later he would be sharing a London stage with British celebrities and comedians, seeking political reforms to defend freedom of expression.

Once in D.C., Wilmschurst spoke at a major cardiology meeting. His topic: the results of a clinical trial to test his hypothesis that a structural heart defect called a patent foramen ovale (PFO) causes a severe type of migraine. The trial, for which Wilmschurst had originally been the senior cardiologist, involved the use of a device called STARFlex to seal PFOs without open heart surgery. It appeared to show that PFO closure did not offer significant benefit to migraine sufferers.

But during his talk, Wilmschurst, who had withdrawn from the trial, alleged that STARFlex's manufacturer—a U.S. medical devices company called NMT Medical, which had funded the trial—had presented the results in a way that masked a disturbing possibility: that the migraines hadn't stopped because STARFlex had failed to close the PFO properly in about one-third of cases. PFOs are principally closed to prevent blood clots from crossing the heart, so failure to do so could prove fatal.

Wilmschurst's comments were picked up by Shelley Wood, a reporter for the specialist cardiology Web site Heartwire based in Montreal, Canada. After the meeting, Wood pub-

lished an article on the STARFlex trial that included quotes from Wilmschurst and others. In December 2007, NMT sued Wilmschurst in the United Kingdom for libel, alleging that he had made claims about STARFlex that he knew were false to prevent his migraine hypothesis from being discredited.

Because of that lawsuit, Wilmschurst has become part of a raging debate in the United Kingdom about whether the nation's libel laws should be reformed. In March, the cardiologist appeared with some of Britain's leading comedians at the Big Libel Gig, an event with the slogan "England's libel laws have become a dangerous joke." Also on stage was journalist Simon Singh, who was then being sued in the United Kingdom for his use of the word "bogus" in a newspaper column about claims made by the British Chiropractic Association. Another recent libel action against Danish radiologist Henrik Thomsen prompted the scientist to vow never to give lectures in the United Kingdom again. In a fourth case involving science, a threat of a defamation suit by the Israeli company Nemesysco Ltd. caused a United Kingdom publisher to pull from its Web site a published paper titled "Charlatanry in forensic speech science" that challenged the science behind the company's lie-detector test. The publisher's action was taken despite the objections of the paper's Swedish authors (*Science*, 13 February 2009, p. 863).

These cases have prompted prominent researchers and U.K. groups such as Index

Staged protest. In London, celebrities such as Dara O'Briain (left) and Ariane Sherine (right) participated in the Big Libel Gig, which was prompted by recent British lawsuits, including ones against scientists and science writer Simon Singh.

on Censorship and Sense about Science to complain that U.K. libel laws—and the high costs of defending a libel action—are forcing researchers and scientific journals to censor or edit academic material. So far, there's little evidence that scientific publishers have been seriously affected: None of 22 journals or journal publishers contacted by *Science* has rejected a research paper solely because of libel concerns, for example. But the publicity surrounding the scientists being sued and several other recent cases involving nonresearchers has led the U.K. government to take a hard look at the nation's libel laws.

Last year, then–Justice Secretary Jack Straw set up a libel working group whose members included the head of the U.K.'s Medical Research Council. The group's report, delivered at the end of March, recommended major changes to the U.K.'s libel laws, several of which could give researchers more protection from lawsuits. Voters in the 6 May U.K. election kicked out Straw's Labour Party, however. The new Conservative–Liberal Democrat ruling coalition has also promised a "review of libel laws to protect freedom of speech," but it is as yet unclear what this will mean in practice.

Libel tourism

Critics of U.K. libel laws point out that, although NMT Medical is an American company, Heartwire is a Canadian Web site, and Wilmschurst's comments were made in the United States, Wilmschurst is being sued in Britain. (Neither Heartwire nor Wood is being sued.) NMT can go after Wilmschurst in the United Kingdom because the Heartwire article, like almost any online publication, can be accessed from computers in the nation, and so the comments could affect the reputation of the company there.

NMT's suit is an example of what has become known as "libel tourism": filing libel suits in countries other than where the alleged offenses occurred. The United Kingdom is viewed by many as a hot spot of libel tourism because the country arguably has some of the most plaintiff-friendly libel laws in the developed world. In the United States, a plaintiff in a libel case must prove that a contested statement is false. If the plaintiff is a figure in the public eye or a corporation, he or she must also prove that the defendant

knew the statement was false or had serious doubts that it was true. Conversely, in the United Kingdom, it is the defendant who bears the burden of proof that a statement is true. In fact, the strong presumption in favor of freedom of speech that exists in U.S. law is found in few other countries; in France and other parts of Europe, libel remains a criminal offense. Still, almost all the recent libel cases involving scientists have been filed in the United Kingdom.

Thomsen learned about the U.K. libel laws the hard way. His unusual case arose from a 15-minute presentation at a radiology conference in Oxford in 2007 in which he detailed his experience at Copenhagen University Hospital with a contrast agent injected to enhance the clarity of MRI scans. Thomsen said at the conference that he noticed that 30 to 40 patients at his hospital with preexisting kidney problems had suffered serious side effects after being injected with the agent, known as Omniscan. Several had been permanently confined to wheelchairs and at least one had died. Thomsen's observations, together with those of other doctors from Denmark and elsewhere in the world, led the U.S. Food and Drug Administration in 2007 to recommend that patients with decreased renal function not be injected with Omniscan or other contrast agents containing the heavy metal gadolinium unless absolutely necessary.

Thomsen's reward was to be sued for libel by GE Healthcare, based in Buckinghamshire, U.K., the manufacturer of Omniscan. In a suit filed in October 2008 in the United Kingdom, the company claimed that Thomsen's Oxford presentation implied that it was knowingly selling a potentially lethal substance and thus his words constituted defamation "by way of innuendo." Earlier this year, in February, GE dropped the suit, 2 days after Thomsen's lawyer told British media that he would file a libel counterclaim against the company. The two parties have now reached a settlement in which GE conceded that Thomsen's presentation did not impute malice to the company's actions. Other settlement details, such as whether GE Healthcare will pay Thomsen's legal costs, have not been made public.

Still, Thomsen has vowed to refuse further speaking engagements in the United Kingdom, where, he says, "it seems you can be

Libel Lawsuits



Defendant: Simon Singh

Plaintiff: British Chiropractic Association

Dispute: BCA objected to a newspaper column by Singh that used the word "bogus" to describe chiropractic claims.

Status: Withdrawn by BCA



Defendant: Peter Wilmshurst

Plaintiff: NMT Medical

Dispute: Firm alleges Wilmshurst knowingly made false claims about a medical device it makes and he tested.

Status: Suit active, no court date set

Defendant: Henrik Thomsen

Plaintiff: GE Healthcare

Dispute: Company claimed Thomsen defamed it by suggesting that it sold a contrast agent called Omniscan knowing it was dangerous.

Status: Withdrawn by GE



sued for telling the truth"—a promise he intends to stand by as long as current U.K. libel laws remain in force.

Costly court cases

Mark Lewis, lawyer for Wilmshurst, argues that U.K. libel laws have a "chilling effect" on research into controversial issues, causing scientist-written articles to be modified or withdrawn in response to a lawsuit threat. Indeed, concern about U.K. laws may have been a factor in a forensic psychology journal article whose publication has been delayed after threat of litigation from another researcher

and a company he founded (see sidebar).

Nemesysco, GE Healthcare, and NMT Medical have all insisted that they simply objected to being accused of dishonest conduct and never intended to stifle scientific debate about the effectiveness or safety of their products. (Nemesysco's letter to Equinox, the publisher of the journal article challenging the basis of the company's lie-detector test, did, however, demand that Equinox admit that the scientific claims in the paper were "largely unfounded"—something Equinox refused to do.) But, citing the Thomsen case as an example, Lewis says, "the point is that whether intended or not, [GE Healthcare] did stifle [scientific debate]."

Thomsen claims that individual researchers are now reluctant to ask provocative questions at academic conferences. "It is my impression that during the last 2 years, the number of questions regarding, for example, contrast agents from the audience in the auditorium is decreasing. [Instead], you get them in the corridors. This means that instead of being able to benefit everyone, communication is just one to one."

Science asked 22 leading scientific and medical journals and journal publishers whether Britain's libel laws are having the chilling effect Lewis claims. The responses varied widely, with some journals having had no experience with libel whatsoever and others having regular contact with a legal team.

For example, American Psychological Association Publisher Gary VandenBos says that prepublication legal threats "are rare, but they occur frequently enough that they are not in any way 'extraordinary' to me, although they would be to the average scholarly author. [In 25 years] I have had to deal [with] about 20 to 30 threats of lawsuits related to manuscripts in prepublication status."

Several journals reported rejecting papers that were "clearly libelous" or removing material that might have attracted libel action, but they insisted that this had been done on editorial grounds. VandenBos notes that APA has published all articles challenged so far, although some may have been legally evaluated and modified. Another publisher noted: "We know that on occasion academics may make assertions they can't then support; the legal advice we obtain often helps them to clarify their thinking and so we end up with a better paper, as well as one that should stand up in any court." No scientific journals reported removing

material on legal grounds that they would not have removed on academic grounds.

The medical journals reported more concerns: *The Lancet*, for example, sends editorial material for review by publisher Elsevier's in-house legal team "about once a month." Compared with commentaries or reviews, research papers are far less likely to require checking by libel lawyers, says *British Medical Journal* Editor Fiona Godlee, although she notes that "even a research paper can have a discussion section" that could raise problems.

Godlee, who has been an outspoken proponent of reform of U.K. libel laws, says science journals need legal advisers—but not all get such advice or can afford it. "I think to make sure you get your story right is exactly what the very best kind of lawyer will help you to do. But then we're a well-resourced journal," she says. "I think the risk is for the individuals who want to speak out or the smaller bodies who may well find themselves unwilling to do so."

One reason academics and some publishers may be nervous is that defending a libel action in the United Kingdom is usually extremely expensive. According to a 2008 report by the University of Oxford's Centre for Socio-Legal Studies, those costs are 140 times the European average and can often exceed £1 million (about \$1.5 million). Singh estimates he has spent £200,000 of his own money to defend his libel action against the British Chiropractic Association. Although the association dropped the suit on 15 April, after an appeals court ruling gave Singh greater leeway in defending his comments, there has been no clear statement by the parties about how much of those costs the association will cover, apparently because that's still in negotiation. "However well this process goes, Simon is likely to be out of pocket by about £20,000," Singh's solicitor Robert Dougans estimates.

Even when a defendant wins a libel action and is awarded costs, the amount rarely covers all legal bills. And the difference can be higher than the claimed damages. Wilmshurst's lawyer says this is a strong incentive for those accused of libel to settle rather than attempt to clear their name: "The general rule of thumb is that you recover two-thirds of your costs," says Lewis. "If a claim costs £500,000 to defend, winning and recovering still costs £166,000—whereas giving in straightaway might only cost £25,000—so there is a cost of sticking up for your right to say something."

Prescription for reform

Attacks on U.K. libel laws are coming from both outside and inside the country. After a Saudi sheik won a libel judgment in the United Kingdom against New York-based criminologist Rachel Ehrenfeld over a book that had never been officially published in Britain, the New York legislature in February 2008 passed the Libel Terrorism Protection Act. (The act's name came from the argument that such libel



"The working group very strongly recommended that a public-interest defense needs to be considered, and that clearly has major implications for scientists."

—LESZEK BORYSIEWICZ,
MEDICAL RESEARCH COUNCIL

cases are an attack on America's freedom of speech.) The law states that the libel judgments of foreign courts are not enforceable in New York unless the laws under which those courts operate are consistent with the First Amendment to the U.S. Constitution.

Online
sciencemag.org
Online Q&A
with Leszek
Borysiewicz.

In the United Kingdom, a coalition of groups called the Campaign for Libel Reform issued a report in November 2009 that contained 10 recommendations. They include placing the burden of proof on plaintiffs to demonstrate that contested statements are false, capping costs and damages, and introducing a statutory public interest defense to protect defendants who shared concerns they had good reason to believe were valid even if they later turned out to be mistaken. A petition for libel law reform launched by the campaign has so far attracted more than 50,000 signatures.

This report prompted Straw to form the working group on libel, a 17-strong committee that included several people from the media and media law (including Thomsen's lawyer Andrew Stephenson), and the directors of the three groups behind the Campaign for Libel Reform. Leszek Borysiewicz, chair of the U.K. Medical Research Council, was the only scientist on the committee, but he says the concerns of scientists were taken seriously: "They certainly made every effort to try to understand the nature of the prob-

lems that [scientists] were experiencing, and throughout they took great care to recognize the issues that scientists have raised."

Some, such as Evan Harris, the Liberal Democrat science spokesperson, have argued that scientists and physicians, or at least their peer-reviewed academic publications, should receive special protections, known as "qualified privilege," that make libel actions more difficult. But the libel

working group didn't endorse that concept in its final report, released on 23 March. "Qualified privilege was discussed," explains Borysiewicz, "but if you start allowing qualified privilege for scientific meetings, at what point does that meeting stop being scientific, and when does qualified privilege stop?"

The report recorded more points of disagreement than agreement among its members, which is perhaps not surprising given the varied backgrounds of the committee members. There was, however, some consensus among the working group on the need to codify a strong public interest defense so that publishers, including those producing science journals, can more safely predict the results of legal challenges and need not fear fighting claims unnecessarily. "That clearly has major implications for scientists," says Borysiewicz. The working group was not asked to examine the issue of costs in detail, but the report also noted that "it is important to ensure that costs of libel proceedings are minimised so far as possible."

May's U.K. government switch has left the future of libel reforms uncertain. The only new libel legislation introduced since the working group report came in late May when Anthony Lester, a member of the U.K. House of Lords, submitted a draft bill offering amendments to the existing libel law that would clarify the public interest defense, as well as shorten court proceedings and reduce their cost; a judge would hear cases, for example, unless there are compelling reasons a jury is needed. Lester also proposes that plaintiffs should, instead of justifying a libel case based on potential harm, have to show that a statement has caused or is likely to cause substantial actual harm and, in the case of a corporation, financial loss.

Others bills or amendments may be put forth, however. "It now becomes even more important that we actually keep a close eye on the reforms that are being proposed because the devil, as always, will be in the detail," says Borysiewicz. —TIM WOGAN

After Legal Threat and 3-Year Delay, Paper on Psychopathy to Appear—Maybe

Have fears of “libel tourism” (see main text, p. 1348) kept a controversial psychology paper on psychopaths from seeing the light of day for more than 3 years after its acceptance by a journal? One of the authors believes that’s the case, citing a 2007 threat of a defamation lawsuit by forensic psychologist Robert Hare of the University of British Columbia, Vancouver. “One of the major issues in this case was that the U.K. (where my coauthor resides) and Canada (where Dr. Hare resides) have much more lax libel laws than those in the U.S. (where I reside),” Jennifer Skeem, a forensic psychologist at the University of California, Irvine, writes in an e-mail to *Science*. “My understanding was that Hare could file anywhere he wished. My coauthor’s university in Glasgow was more concerned than mine, but [agreed to the article going] ahead after months and months of deliberation.”

The disputed paper, which addresses Hare’s views on how to diagnose psychopathy and whether past criminal behavior is central to the condition, will be published in the June issue of *Psychological Assessment*, according to its publisher, the American Psychological Association (APA). Hare argues that he wasn’t responsible for the extended delay, saying he agreed in 2008 to write a response for publication at the same time. Indeed, he’s “frustrated” because the original paper was widely circulated, and he hasn’t had an opportunity to correct what he considers major flaws and misrepresentations of his writings on the connections between criminality and psychopathy. “In 40 years I’ve engaged in some heated academic debates, but in the journals. This time was different because it had little to do with scientific/academic discourse, and everything with taking unwarranted shots at me as a straw man,” Hare says.

Back in 2006, Skeem and psychologist David Cooke of Glasgow Caledonian University submitted a paper to *Psychological Assessment* entitled “Is Criminal Behavior a Central Component of Psychopathy? Conceptual Directions for Resolving the Debate.” Much of the article centers on criticism of the clinical and research tool Hare developed, known as the Psychopathy Checklist-Revised (PCL-R), which is used worldwide to diagnose psychopathy and whose results have been shown to predict future violent behavior (*Science*, 5 September 2008, p. 1284).

PCL-R evaluations are often considered by judges handing out criminal sentences and parole officers debating whether to free prisoners. Hare has founded a company to train people in the use of PCL-R and is paid a royalty when the official checklist is used, but he says the income generated is “modest” and far less than he could earn by testifying as an expert witness, as many in his field do. Hare says he discloses his conflict of interest, but adds that others in his field, including Cooke, who is developing an alternative measure of psychopathy, also have possible conflicts.

Psychological Assessment peer reviewed, accepted, and scheduled Skeem and Cooke’s article for publication in September 2007, but a colleague provided a copy to Hare before then. Hare says he demanded revisions and simultaneous publication of a response. The quarterly journal agreed to a published response, but according to Hare, only “cosmetic” changes were made to the original manuscript. After consulting lawyers who indicated the article could meet Can-

and his company would “have no choice but to seek financial damages from your publication and from the authors of the article, as well as a public retraction.”

Skeem and Cooke have rejected the charges in the e-mail. After the legal threat, APA conducted another review of their paper and asked the authors for additional changes. “We have pored through the entire manuscript again ... and have found no wrong quotes or citations,” Skeem responded in a 2008 letter to APA. In that letter, Skeem and Cooke agreed to new revisions but demanded that APA publish the paper.

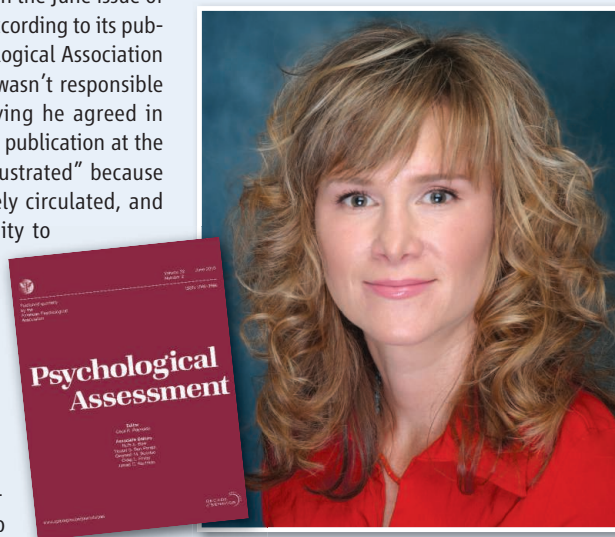
This fracas has only now come to light after psychologist Norman Poythress and attorney John P. Petrila, both of the University of South Florida (USF) in Tampa, noted it in a commentary published online on 13 May in the *International Journal of Forensic Mental Health*. They wrote that the litigation threat had impeded psychopathy research and could allow lawyers challenging PCL-R to claim that its flaws were being suppressed. They further asserted that such threats “strike at the heart of the peer review process, may have a chilling effect on the values at the core of academic freedom, and may potentially impede the scientific testing of various theories, models and products.”

Hare counters that he expected and hoped that the original paper and his response would be published in 2008. The lawsuit threat was meant only to get the “attention” of APA, Skeem, and Cooke and force changes to the article. “We haven’t considered legal action in over two years,” he says. Hare says he’s upset colleagues are suggesting he squelched debate and says the lawsuit threat may not have been “judicious.” “Maybe I went off half-cocked.”

Hare and Skeem both criticize APA about the long publishing delay. Asked about the episode, APA Publisher Gary VandenBos comments: “APA has never refused to publish an article because of a threatened defamation lawsuit, but if confronted with a claim of defamation before publication, APA would, of course, take reasonable steps to assure that the material APA is about to publish is accurate.” The APA journal will now publish Skeem and Cooke’s paper, a response from Hare and a colleague, and a reply from Skeem and Cooke, says VandenBos. “Nothing involved in this involves a legal negotiation or legal agreement with Dr. Hare,” he adds.

Hare and Skeem both remain skeptical. “I am not confident the papers will be published this month,” says Skeem. “I’m still not convinced these articles will come out,” says Hare. At least they agree on something.

—JOHN TRAVIS



Time out. *Psychological Assessment* is planning to publish a paper by Jennifer Skeem and a co-author, along with a critical response, 3 years after the paper was first accepted.

ada’s definition of defamation, Hare had legal counsel send a letter to APA and the authors demanding that the publication be pulled.

According to an e-mail *Science* has obtained, Hare’s lawyers wrote on 8 November 2007 that “publication of the article would be detrimental to Dr. Hare professionally and financially” and that “it is our position that the authors deliberately fabricated or altered quotes of Dr. Hare.” If publication went ahead, the e-mail said, Hare

Sexual selection

1356

Temperature control
in ancient reptiles

1361

Chloride
exchangers

1364



LETTERS | BOOKS | POLICY FORUM | EDUCATION FORUM | PERSPECTIVES

LETTERS

edited by Jennifer Sills

The Status of Atlantic Bluefin Tuna

WE ARE CONCERNED BY SOME STATEMENTS IN THE LETTER "THE FATE of Atlantic bluefin tuna" by tuna scientist J.-M. Fromentin (12 March, p. 1325). Fromentin's comments reflect widespread misconceptions used to argue against the inclusion of Atlantic bluefin tuna (BFT) in Appendix I of the Convention on International Trade in Endangered Species (CITES).

Fromentin's statement that current scientific knowledge does not unequivocally support a BFT listing under Appendix I doesn't serve the debate. Very rarely does fisheries science unequivocally support any conclusion; rather, it expresses a high probability of a given scenario being met. Fromentin omits the fact that the International Commission for the Conservation of Atlantic Tunas (ICCAT) Scientific Committee concluded (1) that there was a 95% probability that BFT had declined to the extent that it would qualify for an Appendix I listing. This conclusion was endorsed by the majority of the FAO Panel (2), the International Union for Conservation of Nature (IUCN) (3), and the CITES Secretariat (4).

Although CITES decided not to list the Atlantic BFT on Appendix I, it is not true that it would have been the first commercially exploited marine species so listed. Most of the great whales, all marine turtles, all but one sawfish species, and two sturgeon species are listed on CITES Appendix I.

We regret that many governments and individuals continue to be more concerned about future CITES listings of other marine species than about the very reasons why those may be eligible: the dire state of many fish stocks. Nothing indicates, as Fromentin suggests, that the

listing of BFT could lead to CITES implicitly assuming management of fisheries. Such a statement is absurd. CITES deals specifically with international trade only, and its mandate is therefore completely different from fisheries organizations. In cases where international trade drives overfishing, fisheries organizations and CITES working together offers the best possible chance of ensuring sustainability.

SEBASTIAN LOSADA,^{1*} SUSAN LIEBERMAN,²
CARLOS DREWS,³ MICHAEL HIRSHFIELD⁴

¹Greenpeace International, San Bernardo, 107, Madrid 28015, Spain. ²The Pew Environment Group, 901 E Street, NW, Washington, DC 20004, USA. ³WWF International, Avenue du Mont Blanc, Gland 1196, Switzerland. ⁴Oceana, 1350 Connecticut Avenue, NW, Washington, DC 20036, USA.

*To whom correspondence should be addressed. E-mail: slosada@greenpeace.org

References

1. ICCAT, "Extension of the 2009 SCRS meeting to consider the status of Atlantic bluefin tuna populations with respect to CITES biological listing criteria" (ICCAT, Madrid, 2009).
2. FAO, *Report of the Third FAO Expert Advisory Panel for the Assessment of Proposals to Amend Appendices I and II of CITES Concerning Commercially-Exploited Aquatic Species* (FAO Fisheries Report No. 925, Rome, 2010).
3. IUCN/TRAFFIC, *Analyses of the Proposals to Amend the CITES Appendices* (Information document CoP15, Inf. 18A, CITES, 2010).
4. CITES, *Comments from the Parties and Comments and Recommendations from the Secretariat* (Working document CoP 15, Doc 68, Annex 2, CITES, 2010).

Response

LOSADA ET AL. EXPRESSED CONCERN ABOUT the level of scientific evidence that supports listing bluefin tuna in Appendix I. It is a fact that there are great uncertainties in determining whether the current stock status of bluefin tuna (BFT) really meets the CITES biological listing criteria. The scientific committee of ICCAT, the group of international experts that is responsible for assessing BFT and had made the original diagnosis of overexploitation, was solicited to perform this analysis (1). Although two CITES criteria were not relevant for BFT, a third one—"a marked decline"—could be. However, evaluating this criterion necessitated the estimation of key parameters that are currently highly uncer-

tain, particularly the reference biomass level. Therefore, two historical baselines were considered: (i) the highest observed historical level and (ii) the virgin biomass (the biomass that would be expected without fishing). There is only a 21 to 30% probability that the former meets the CITES criterion, whereas there is more than a 92% probability in the case of the latter (1). There was no scientific consensus as to which of the two options to choose as the reference level. This critical scientific uncertainty, which was also reported by the FAO panel, was absent from the public debate, as in Losada *et al.*'s response.

Regarding the possible precedent, I should have been more specific. It is correct that BFT would not have been the first commer-

cially exploited marine species to be listed on Appendix I, but it would be the first commercially exploited marine bony fish species to be so listed.

Regarding the role of CITES, I agree that fisheries organizations and CITES could work together (as CITES and FAO do), but this is a challenging task (2). The poor performance of the fisheries organizations has been thoroughly analyzed and transparently reported [e.g., (3)]. It has led to 100 fish species (or more) that are as overfished or even more overfished than BFT (4, 5). Therefore, I reckon that the same process could logically lead to the listing of these 100 species, which raises the question: Would CITES do better? In the specific case of BFT, Appendix I would



Downloaded from www.sciencemag.org on June 10, 2010

ban most of the legal fishing but may have little impact on illegal fishing. As with many CITES-listed species of high value, a lucrative black market is likely to develop. Therefore, the fishing pressure could remain at a substantial, but unknown, level. Meanwhile, the Appendix I listing will also impair future scientific advice because the ban of most legal fishing fleets will curtail the flow of adequate information on which BFT stock assessment is based.

Unfortunately, the debate around BFT has become too political and, thus, too black and white. In such an arena, the scientific advice is truncated, sometimes distorted, while the inherent complexity and uncertainties are simply ignored. Stakeholders and fisheries lobbies have played this game for decades, leading to the severe overfishing of BFT. It is more than timely to have an open-minded and dispassionate debate and crucial to separate science and political issues.

JEAN-MARC FROMENTIN

IFREMER, Centre de Recherche Halieutique Méditerranéenne et Tropicale, avenue Jean Monnet, BP 171, 34203 Sète cedex, France. E-mail: jean.marc.fromentin@ifremer.fr

References

1. ICCAT, "Extension of the 2009 SCRS meeting to consider the status of Atlantic bluefin tuna populations with respect to CITES biological listing criteria" (ICCAT, Madrid, 2009).
2. P. Doukakis *et al.*, *Conserv. Biol.* **23**, 841 (2008).
3. S. Garcia, J. R. Grainger, *Philos. Trans. R. Soc. London Ser. B* **360**, 21 (2005).
4. FAO, *The State of the World Fisheries and Aquaculture 2008* (FAO Fisheries and Aquaculture Department, Rome, 2009).
5. B. Worm *et al.*, *Science* **325**, 578 (2009).

Protect Pharmaceutical Innovation

IN HER POLICY FORUM "FIXING THE LEGAL framework for pharmaceutical research" (26 February, p. 1083, published online 11 February), GlaxoSmithKline attorney S. M. Knowles contends that the Hatch-Waxman Act should be altered to increase data exclusivity from 5 to 14 years. If implemented, this disturbing change would likely narrow the drug development pipeline and certainly increase drug prices.

Knowles reports the average R&D costs per drug in 2008 to be \$1.2 billion. However, when the analysis accounts for tax deductions given to the pharmaceutical industry and removes opportunity costs from the expenses column, the results show an estimate of one-seventh the cost (including drug failures) (1) of Knowles' industry-funded analysis. Knowles claims that the inadequate patent system does not allow for cost recovery, yet pharmaceutical companies have garnered three times the profit of the average

CORRECTIONS AND CLARIFICATIONS

Reports: "Erosion of lizard diversity by climate change and altered thermal niches" by B. Sinervo *et al.* (14 May, p. 894). In note 32, FONDYCE should have been FONDECYT.

Review: "Arguing to learn in science: The role of collaborative, critical discourse" by J. Osborne (23 April, p. 463). The credit for Fig. 2 did not appear in print; it is RYAN MCVAY/PHOTODISC/GETTY IMAGES.

Fortune 500 company over the past decade (2). Knowles also suggests that the federal government should pay for the use of innovator safety and efficacy data by generic competitors, but she fails to acknowledge that worldwide, government and nonprofits already fund the majority of the health R&D that generates these discoveries (3).

Since its inception in 1984, the Hatch-Waxman Act has saved the U.S. \$734 billion by providing an effective conduit for lower-priced, generic pharmacotherapies to come to market. Generic drugs now make up 70% of prescriptions dispensed and account for only 20% of dollars spent on medications in the United States (4). Hatch-Waxman has been preserving the precarious balance between public welfare and industry incentive for drug innovation by providing affordable medicines.

We agree with Knowles on one point: Hatch-Waxman requires reform to widen the R&D pipeline. Congress must ban brand-name pharmaceutical firms' unethical "pay-for-delay" agreements with generic competitors. In doing so, it would create a paradigm shift, disincentivizing compensatory agreements and, instead, refocusing brand-name companies on constructing novel therapies that may address presently unmet medical needs.

LAURA W. MUSSELWHITE^{1*} AND JANE ANDREWS²

¹Duke University School of Medicine, Durham, NC 27710, USA. ²Johns Hopkins School of Medicine and School of Public Health, Baltimore, MD 21205, USA.

*To whom correspondence should be addressed. E-mail: laura.musselwhite@duke.edu

References

1. B. Young, M. Surrusco, "Rx R&D myths: The case against the drug industry's R&D 'scare card,'" *Public Citizen* (July 2001).
2. The Henry J. Kaiser Family Foundation, "Profitability of pharmaceutical manufacturers" (16 September 2009).
3. Global Forum for Health Research, "Monitoring financial flows for health research" (2006).
4. A. B. Engelberg *et al.*, *N. Engl. J. Med.* **361**, 1917 (2009).

Response

MUSSELWHITE AND ANDREWS MAKE THE unsupported argument that increased incentives (through longer exclusivity periods) will lead to decreased innovation or numbers of new pipeline drugs. On the contrary, a National Academies study recommends extending the data exclusivity period to 12

to 14 years to avoid competitive disadvantage in the United States "given the complexity and length of drug development today" [p. 190 in (1); (2)].

Musselwhite and Andrews' statement that generic drugs now make up about 70% of pharmaceutical prescriptions and account for about 20% of consumer spending on drugs strongly supports my conclusion that the rate of new innovator drugs is not keeping pace with the loss of innovator pharmaceutical markets. This trend is confirmed by a recent report of the Congressional Budget Office (3), which documents that introductions of innovative new molecular entities spiked in the mid- to late 1990s and have been decreasing since 2000, in most years, to levels not seen since the mid-1980s. The Congressional Budget Office also notes that production of new molecular entities that provide a "significant therapeutic or health advance" has decreased from an average of 13 per year in the 1990s to 10 per year in the 2000s (3). New drug launches have been correlated with increased longevity and quality of life (4).

Musselwhite and Andrews assert that the true cost of drug discovery and development is about one-seventh of that found by DiMasi, citing an earlier study done by Ralph Nader's Public Citizen group. The Public Citizen publication is non-peer-reviewed economic analysis. The Public Citizen study contends that the actual out-of-pocket cost per drug is between about \$57 and 110 million, including the cost of failures. As such, it concludes that pharma R&D is less risky than it appears. If this were true, common sense tells us that there would be many more drugs on the market today.

The statistics in the Kaiser Permanente report—that pharmaceutical companies have received three times the level of profits of the average Fortune 500 companies over the past 10 years—cannot be clearly interpreted because they lack definitions for inclusion and include confounding issues (5). One unambiguous and objectively measurable metric is share price: Since 2000, the share prices of many of the largest global pharmaceutical companies have declined dramatically, several by roughly 60% or more (6). Unexpected losses of U.S. market exclusivity under Hatch-Waxman have contributed substantially to this downward view of the

sector by the investment community.

Musselwhite and Andrews suggest that government and nonprofits already fund the majority of the R&D that generates pharma innovation. In fact, valuable NIH-funded R&D complements and supports private R&D, typically through pre-innovation basic research—far removed from commercial development and not appropriate for private investment—and certain clinical trials. It is wholly incorrect to suggest that public grants and investments fund the bulk of private pharma R&D. Increased public spending stimulates and in some cases may be necessary, but is not a substitute for private spending (3).

Innovator and generic pharmaceutical companies are required to file all patent litigation settlement agreements with the Federal Trade Commission (FTC) (3). Since 2004, approximately 152 final settlement agreements have been received by the FTC (7). The FTC has chosen to challenge only a very small number of these settlements and, of those challenged, has lost more than it has won in appellate courts before experienced judges [e.g., (8)]. The misnomer “pay-for-delay settlements” does not reflect the reality that the overwhelming number of settlements

are pro-competitive and not objectionable even to the FTC. Furthermore, the settlement rate for pharmaceutical/generic patent litigation is less than that for patent settlements across all industries, which includes medical devices, biopharmaceutical products, consumer healthcare, important electronics, and software (9–12).

SHERRY M. KNOWLES

Chief Patent Counsel, GlaxoSmithKline, King of Prussia, PA 19406, USA. E-mail: sherry.m.knowles@gsk.com

References and Notes

1. Committee on Science, Engineering, and Public Policy, *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (National Academies Press, Washington, DC, 2007), pp. 185–92.
2. The Patient Protection and Affordable Health Care Act of 2010, signed by President Obama on 21 March 2010, provides 12 years of data protection for follow-on biologics.
3. Congressional Budget Office, “Pharmaceutical R&D and the evolving market for prescription drugs” (October 2009).
4. Medicare Prescription Drug Improvement and Modernization Act of 2003, Pub. L. No. 108-173, Title XI, §1112, 117 Stat. 2066 (8 December 2003) (“MMA”).
5. Several sectors in the overall Fortune 500 have had very significant issues that could have dramatically influenced the overall average profit (e.g., AIG, Fannie Mae, Freddie Mac, automotive companies, and airlines). Furthermore, sometimes companies classified too broadly as pharmaceutical companies generate a significant portion of their profits from other health care sectors such as consumer products (OTC medicines, oral care products), medical imaging, diagnostics testing, and

medical devices such as heart stents.

6. Cumulative Stock Price Performance from 1 January 2000 through 28 February 2010 collected from Bloomberg Financial News: Pfizer –62%, Merck –58%, BMS –65%, and Eli Lilly –63%.
7. Federal Trade Commission, Pharmaceutical Agreement Filings (www.ftc.gov/bc/healthcare/drug/index.htm).
8. *Schering-Plough Corp. v. FTC*, 402 F.3d 1056 (11th Cir. 2005).
9. Of 370 Hatch-Waxman cases studied in 2000–2010: 47% settled, 7% were dropped, the generic lost 24%, and the innovator lost 22%.
10. RBC Capital Markets Report, “Pharmaceutical-analyzing litigation success rates,” 15 January 2010.
11. J. P. Kesan, G. G. Ball, *Wash. U. Law. Rev.* **237**, 259 (2006).
12. The Health Care Reform Bill, H.R. 3590, passed by the House of Representatives on March 22, 2010, correctly omitted the proposed Hatch-Waxman patent settlements provision, based on S. 369 introduced by Senator Kohl on February 2009.

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 3 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.



Learn how current events are impacting your work.

ScienceInsider, the new policy blog from the journal *Science*, is your source for breaking news and instant analysis from the nexus of politics and science.

Produced by an international team of science journalists, *ScienceInsider* offers hard-hitting coverage on a range of issues including climate change, bioterrorism, research funding, and more.

Before research happens at the bench, science policy is formulated in the halls of government. Make sure you understand how current events are impacting your work. Read *ScienceInsider* today.

www.ScienceInsider.org

ScienceInsider

Breaking news and analysis from the world of science policy



HISTORY OF SCIENCE

Looking at Sexual Selection

Karen A. Rader

Sex never really goes away—but the role of female choice in mating has been a shifting object of social and scientific attention. In 1990, Camille Paglia famously celebrated singer Madonna as a prosex, pro-woman feminist: “she taught young women to be fully female and sexual while still exercising total control over their lives” (1). Yet 15 years (and many heated exchanges with other academic feminists) later, Paglia skewered the same pop icon for “still hammering at sex” and refusing to age gracefully (2). Erika Milam’s *Looking for a Few Good Males: Female Choice in Evolutionary Biology* reminds us that scientists have been, in many ways, no less fickle. Biologists over the past century and a half have had a changing relationship with the problems and possibilities posed by courtship behavior in animals.

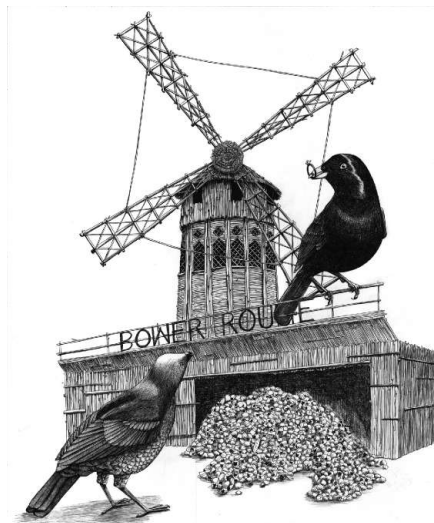
Female choice, Milam notes, has “a history we thought we knew.” It starts with Charles Darwin’s 1871 discussion [in *Descent of Man* (3)] of aesthetics and sexuality: Male ornamentation, Darwin argued, was the result of females comparing male mating displays and choosing the most appealing mate. Along with male-male competition, female choice accounted for varieties within a species. But in the mid-20th century, sexual selection was “eclipsed” by natural selection (largely as a result of the successes of classical and mathematical population geneticists)—until its rebirth in the 1970s, when the study of animal behavior was reunited with evolutionary theory by Robert Trivers and (later) sociobiology. However, this standard history reveals little of how understandings of female choice have shaped—and been shaped by—ideas about the relation between human and animal minds and by changing theories and practices in biology more broadly. Alternatively, Milam uses the topic of female choice as a lens through which to view intellectual, disciplinary, and social developments in the life sciences.

Milam, a University of Maryland historian with a background in biology, organizes her account around a set of questions that researchers asked themselves about when and how animal behavior could explain human courtship interactions. Darwin’s initial debates with Alfred Russel Wallace about

whether female choice was a viable mechanism of evolutionary change turned on disagreements about the psychological continuity of human and animal minds.

Wallace at first rejected sexual selection by female choice because he did not believe animals (regardless of gender) possessed the ability to reason. Some early feminists (such as Charlotte Perkins Gilman) saw potential in sexual selection for establishing the inferior place of women in Victorian society as unnatural. But few scientists engaged this cultural debate until the Progressive era, when American and British biologists reframed both animal courtship (as “an evolutionary forerunner to marriage choice in humans”) and animal copulation (as emotionally and physiologically akin to human love-making). Starting in the 1920s, biologists from Ronald Fisher to Julian Huxley grappled with whether sexual selection in animals countered or enhanced the effects of natural selection. Meanwhile, popularizers garnered attention for their warnings that human mate selection might be unfairly influenced by mere aesthetic concerns, rather than eugenically sound ones (such as high intelligence).

Biologists’ conceptions of female choice became even more complex in the 1930s,



Female choice and male gift-giving. Satin bowerbirds (*Ptilonorhynchus violaceus*).

Looking for a Few Good Males
Female Choice in Evolutionary Biology

by Erika Lorraine Milam

Johns Hopkins University Press,
Baltimore, 2010. 246 pp. \$60,
£31.50. ISBN 9780801894190.

when naturalist and experimentalist traditions converged at places like the American Museum of Natural History. Herpetology curator Gladwyn Kingsley Noble investigated the mating behavior of a wide variety of animals, from salamanders to box turtles, in the AMNH Laboratory of Experimental Biology, where he blended laboratory and field methods. Noble brought anoles into the museum and designed a series of experiments showing how females chose males on the basis of vigorous male displays of their red dewlaps. But when “data gathered in nature differed from his laboratory results, Noble made the field observations trump.” Thus, when he observed homosexual interactions between anoles in the museum’s cages, he attributed it to lack of space and expanded their greenhouse facilities to “correct” the problem. Noble ultimately hoped to construct an evolutionary map of sexual behavior in animals “from fish to man.” His successors, in turn, studied females’ ability to distinguish between males of their own species and males of other species, because they were interested in animal behavior as a mechanism for (rather than a result of) evolutionary change.

After World War II, female choice and sexual selection were marginalized by the research programs of many British animal behavior biologists: their mechanistic explanatory framework just couldn’t accommodate such concepts. Work in ethology (by Nikolaas Tinbergen, Konrad Lorenz, and others) on male-male aggression “reconfigured the functional causes of courtship behavior without recourse to conscious choice on the part of their animal subjects.” The notion of ritualized mating behaviors, however, became controversial when transferred in popular accounts from bowerbirds to people. Later, population geneticists (some of whom were women) described the “rare male” mating preference in female *Drosophila*. Milam suggests that these scientists faced less resistance because they were focused on experimentally and mathematically demonstrating how female choice acted to maintain genetic diversity within a population—but even they believed “flies stood in for humans.”

In the wake of Trivers’s claim that female choice could be explained by differential parental investment (females who mated less than males and had primary caregiving responsibilities would be inherently more choosy), organismal biologists proudly reclaimed sexual selection as an experimental field science in the 1970s. Work on

The reviewer is at the Science, Technology, and Society Program, Virginia Commonwealth University, 916 West Franklin Street, Room 201, Richmond, VA 23284–2001, USA. E-mail: karader@vcu.edu

sexual selection “became entangled within ... disciplinary divides” at a moment when animal behavior researchers sought to establish their authority and novelty in the molecular age. Sociobiological accounts ensured more popular and scientific attention in the 1980s: female choice was a bitter pill to swallow when reframed in stereotypes of coy females luring eager males.

Looking for a Few Good Males, unlike some earlier, more philosophical accounts (4), is not exclusively about female choice.

But that, in fact, is the point: in Milam's view, “the differentiated histories of female choice and sexual selection illustrate how evolutionary theory and animal behavior were continuously resynthesized throughout the twentieth century.” Sensitive to the conceptual nuances and disciplinary politics undergirding this process, Milam provides an invaluable synthesis for historians of biology, scientists, and those with a popular interest in animal studies. Female choice, it seems, is a scientific problem that, like the iconic pop star, also

refuses to age gracefully—and Milam would have it no other way.

References

1. C. Paglia, *New York Times*, 14 December 1990, p. A39.
2. www.salon.com/entertainment/feature/2005/12/02/madonna.
3. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (John Murray, London, 1871).
4. H. Cronin, *The Ant and the Peacock: Altruism and Sexual Selection from Darwin to Today* (Cambridge Univ. Press, Cambridge, 1991).

10.1126/science.1189754

HUMAN FACTORS

We Should All Have a Little List

Michael A. Goldman

Medicine, construction, and flying a jet have one thing in common. They've all become very complex endeavors in which simple mistakes can cost lives.

When we look closely, we recognize the same balls being dropped over and over, even by those of great ability and determination.... It's time to try something else.

Try a checklist.

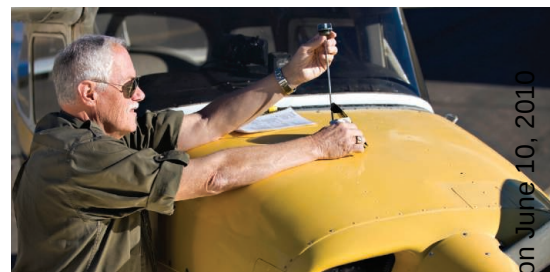
So exhorts Harvard Medical School surgeon Atul Gawande in *The Checklist Manifesto: How to Get Things Right*.

The value of checklists in aviation and space flight is legion. When a B-17 prototype crashed and burned in 1935, after a pilot “had forgotten to release a new locking mechanism on the elevator and rudder controls.... Boeing nearly went bankrupt.” But a group of test pilots proposed a low-tech solution—a checklist to make sure that didn't happen again. The B-17 went on to become a safe and successful aircraft, with more than 13,000 sold to the army. Checklists have since proliferated in aviation, with subtle reminders for what to do in various emergencies as well as in routine flight. When things go really wrong, as when the encounter of US Airways Flight 1549 with a flock of geese shut down both engines, “[t]he crew ... showed an ability to adhere to vital procedures when it mattered most, to remain

calm under pressure, to recognize where one needed to improvise and where one needed *not* to improvise.” Gawande comments, “This is what it means to be a hero in the modern era. These are the rare qualities that we must understand are needed in the larger world.”

Make no mistake about it. Here—and elsewhere (*1*)—Gawande makes it abundantly clear that checklists can make being a patient in surgery much safer and can do so just about anywhere in the world. Surgical complications in eight study hospitals fell 36%, deaths by 47%, with 4000 patients experiencing 277 complications instead of the expected 435. It is not surprising that, as Gawande points out, even though 20% of medical professionals using checklists don't like doing so and didn't think they were useful, 93% would want one used if they were on the operating table. What is surprising is that the use of checklists in surgery has not yet fully penetrated the profession. A lot of what healthcare reform must accomplish to reduce costs is the reduction of errors in medicine, and one might expect pressure to mount.

A checklist as an essential tool for a professional, skilled worker, or knowledge worker is counterintuitive and to some, downright insulting. Unfortunately, this is the idea that projects from the book's cover and in nearly every account of it I've seen. But I think two other points are at least as important aspects of Gawande's thesis. These involve the ability to communicate and work as a team and the ability to handle the unexpected. Both capabilities, according to Gawande, are enhanced rather than stifled by the regimentation of a checklist. The safe surgery checklist that



Pilot's preflight. Crucial checklist.

Gawande helped develop (*1*) instructs the entire team to introduce themselves by name and role and to discuss the unusual aspects of the case and potential problems. The checklist distributes power, so that a nurse reading a checklist acquires the authority, as a member of the team, to stop the surgeon from omitting a critical step or making a stupid mistake. Gawande presents numerous examples of problems anticipated just because everyone together went through the checklist, communicated, and was ready for the unforeseen. He credits checklists with freeing the pilot and co-pilot on US Airways Flight 1549 to make the critical decisions without being bothered with the routine.

I would like to see more about the utility of checklists in a larger context. Gawande illustrates the utility of checklists in the construction of modern skyscrapers, in complex decisions in investing, and even in a restaurant kitchen. While Gawande's treatise makes the case for these, it leaves me wanting to know much more about how to apply his research in the complex daily life and work of the readers of *Science*.

In all, *The Checklist Manifesto* is an informative, entertaining, and even inspiring read. But I'm still glad to check it off my reading list!

References

1. A. B. Haynes *et al.*, *N. Engl. J. Med.* **360**, 491 (2009).

10.1126/science.1190165

The Checklist Manifesto How to Get Things Right

by Atul Gawande

Metropolitan Books
(Henry Holt), New York, 2010.
221 pp. \$24.50, C\$29.50.
ISBN 9780805091748.

The reviewer is at the Department of Biology, San Francisco State University, San Francisco, CA 94132-1722, USA. E-mail: goldman@sfsu.edu

Biodiversity Conservation Research, Training, and Policy in São Paulo

The BIOTA-FAPESP program is linking a decade of research on biodiversity into public policy in the state of São Paulo.

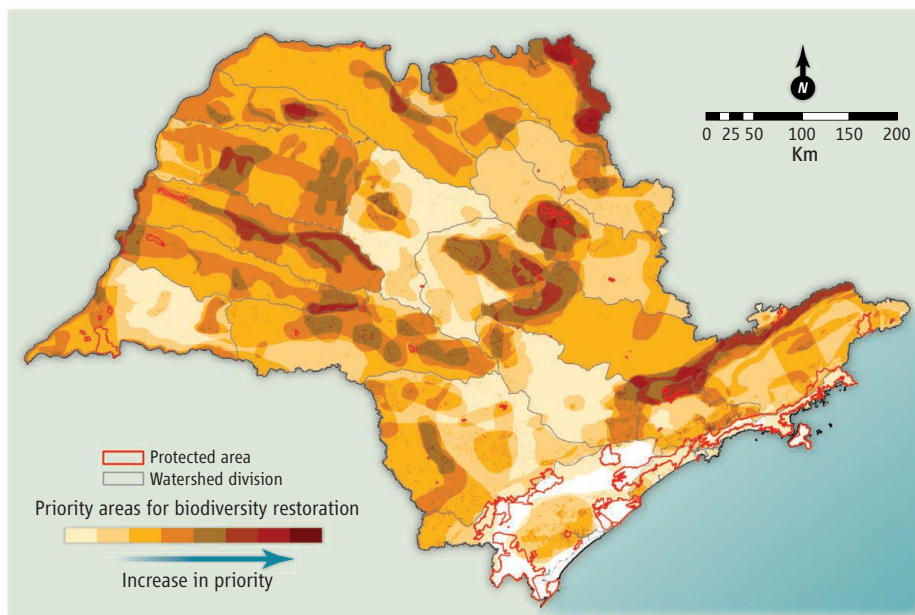
Carlos A. Joly,^{1*} Ricardo R. Rodrigues,² Jean Paul Metzger,³ Célio F. B. Haddad,⁴ Luciano M. Verdade,² Mariana C. Oliveira,⁵ Vanderlan S. Bolzani⁶

Since the Convention on Biological Diversity (CBD) in 1992, biodiversity conservation (the protection of species, ecosystems, and ecological processes) and restoration (recovery of degraded ecosystems) have been high priorities for many countries. Scarce financial resources must be optimized, especially in developing countries considered megadiverse (1), by investing in programs that combine biodiversity research, personnel training, and public-policy impact. We describe an ongoing program in the state of São Paulo, Brazil, that may be a useful example of how conservation initiatives with a solid scientific basis can be achieved.

São Paulo's rich native biodiversity is threatened by changes in land cover and fragmentation (2, 3). This prompted scientists in 1999 to found the Virtual Institute of Biodiversity, BIOTA-FAPESP. FAPESP, the State of São Paulo Research Foundation, is a nonpolitical, taxpayer-funded foundation, one of the main funding agencies for scientific and technological research in Brazil, and a supporter of this program.

The program's scope of research ranges from DNA bar-coding to landscape ecology and includes taxonomy, phylogeny, and phytogeography, as well as human dimensions of biodiversity conservation, restoration, and sustainable use. During its first 10 years, the program supported 94 major research projects, described more than 1800 new species, acquired and archived information on over 12,000 species, and made data from 35 major biological collections available online, a first for Brazilian biological collections.

In 2001, the program launched an open-access, electronic, peer-reviewed journal, *Biota Neotropica* (4), to publish research



Priority areas for biodiversity restoration in São Paulo. The figure also shows the existing network of state parks (red lines) and the state's division of Water Management Units (gray lines). (See SOM.)

results on biodiversity in the Neotropics. In 2002, the program began *BIOprospecTA*, a venture to search for new bioactive compounds of economic interest that has already resulted in three prototype patents.

Policy Impact

Between 2006 and 2008, BIOTA-FAPESP researchers made a concerted effort to synthesize data for use in public-policy-making. Scientists worked with the state secretary of the environment and nongovernmental organizations (NGOs) such as Conservation International, The Nature Conservancy, and the World Wildlife Fund. The synthesis was based on more than 151,000 records of 9405 species (table S1), as well as landscape structural parameters and biological indices from over 92,000 fragments of native vegetation. Two synthesis maps, identifying priority areas for restoration (see the figure, above) and conservation (fig. S1), together with other detailed data and guidelines (5), have been adopted by São Paulo state as the legal framework for improving public policies on biodiversity conservation and restoration, such as prioritizing areas for

forest restoration (as one means of reconnecting fragments of native vegetation) and selecting areas for new Conservation Units. There are four governmental decrees and 11 resolutions [see supporting online material (SOM)] that quote the BIOTA-FAPESP guidelines. Before this effort was made, most policy decisions were based on secondary data of heterogeneous quality, not evaluated by a scientific committee.

One of the most striking implementations of BIOTA-FAPESP recommendations is a joint resolution of the state secretaries of the environment and of agriculture to establish an agro-ecological zoning ordinance that prohibits sugarcane expansion to areas that are priorities for biodiversity conservation and restoration (fig. S2). Acceptance of these recommendations may be linked to commercial demands from the international ethanol market, which is increasingly requiring compliance with environmentally sound commodity production practices.

This experience provides an example for other regions. Maps showing priority areas for biodiversity restoration have been produced for the entire area originally covered

¹Department of Plant Biology, Biology Institute, State University of Campinas, 13083-970 Campinas, São Paulo (SP), Brazil. ²Department of Biological Science, Luiz de Queiroz College of Agriculture, State University of São Paulo, Piracicaba, SP, Brazil. ³Department of Ecology, Institute of Biosciences, University of São Paulo, São Paulo, SP, Brazil.

⁴Department of Zoology, Institute of Biosciences, Universidade Estadual Paulista, Rio Claro, SP, Brazil. ⁵Department of Botany, Institute of Biosciences, University of São Paulo, São Paulo, SP, Brazil. ⁶Department of Organic Chemistry, University Estadual Paulista, Araraquara, SP, Brazil.

*Author for correspondence. E-mail: cjoly@unicamp.br

by the Atlantic forest in 17 Brazilian states (6). Other Brazilian states have begun long-term programs based on the BIOTA-FAPESP guidelines. The Brazilian National Research Council (CNPq) is planning a similar initiative and, likewise, the U.S. National Science Foundation recently launched the program, Dimensions of Biodiversity.

Keys to Success

What makes a program on biodiversity conservation simultaneously successful in research, training, and policy (7)? Several external factors may have contributed to progress thus far: a consolidated network of research institutions, graduate programs, and biodiversity researchers in the state of São Paulo; pressure from commodity markets for certification; increasing social awareness of biodiversity conservation and demand for scientifically sound policies; the large network of 64 state parks and reserves; the political will demonstrated by the state secretary of the environment in supporting the program. Political and economic stability in Brazil were also important factors that allowed FAPESP to make a crucial, long-term (10-year) commitment to funding, providing an average annual budget of U.S.\$2.5M.

But particular aspects of the program must also be recognized as important. It is a

research-driven initiative—planned, implemented, and coordinated by scientists—in contrast with most previous Brazilian conservation policies. The funding agency, FAPESP, has de facto political and administrative autonomy, which allows it to invest in long-term scientific programs and to ensure quality through a rigid peer-review standard, which is rare in Brazil. The program is evaluated by an international committee every 2 years (8). Members of the committee represent diverse areas of scientific expertise; one of the members was from the senior administration of the secretary of the environment, which helped bridge the gap between scientists and policy-makers. The fact that the program is fully based on the CBD, which provides an undisputed legal framework, is another crucial factor.

Strong ties with collaborators are also essential. Many technical staff of both the state and the NGOs developed student projects, supported by the BIOTA-FAPESP program, and became strong allies in producing the synthesis and implementing biodiversity conservation and restoration priorities. Further research by independent, external evaluators is needed.

The program has not been successful in all areas. It failed to translate scientific advancements into teaching material for use in schools. It did not study the entire state

well enough to establish priority areas in every watershed, and the marine ecosystems were not studied in the same depth as continental ones. Also, the present distribution and risk of invasive species have not been mapped, and few projects focused on the human dimensions of biodiversity conservation. These gaps were identified during internal and external evaluation in 2009, and are thus priorities in the *Science Plan and Strategies for the Next Decade* (see SOM).

References and Notes

1. C. R. Margules, R. L. Pressey, *Nature* **405**, 243 (2000).
2. M. C. Ribeiro, J. P. Metzger, A. C. Martensen, F. J. Ponzoni, M. M. Hirota, *Biol. Conserv.* **142**, 1141 (2009).
3. G. Durigan, J. A. Ratter, *Edinb. J. Bot.* **63**, 119 (2006).
4. *Biota Neotropica*, www.biotaneotropica.org.br.
5. R. R. Rodrigues et al., *Diretrizes para a Conservação e Restauração da Biodiversidade no Estado de São Paulo* (Botanical Institute, BIOTA-FAPESP, São Paulo, Brazil, 2008); www.ambiente.sp.gov.br/cpla/files/100111_biota_fapesp.pdf.
6. Pacto pela Restauração da Mata Atlântica, www.pacto.org.br/pdf/mapeamento-areas-restauracao.pdf.
7. P. Shanley, C. López, *Biotropica* **41**, 535 (2009).
8. Biota, www.biota.org.br/info/sac/.
9. Supported by FAPESP and CNPq. We thank M. C. Ribeiro, M. Keller, and especially M. C. W. de Brito and V. L. R. Bononi of the state of São Paulo secretary of the environment, who provided financial and technical support for workshops organized to produce the synthesis maps.

Supplementary Online Material

www.sciencemag.org/cgi/content/full/328/5984/1358/DC1

10.1126/science.1188639

PUBLIC HEALTH

Global HIV/AIDS Policy in Transition

John Bongaarts,^{1*} and Mead Over²

In 2007, the United Nations Joint Programme on HIV/AIDS (UNAIDS) concluded that “Global HIV incidence likely peaked in the late 1990s” (1), due to “natural trends in the epidemic as well as the result of prevention programmes” (1). The slow decline in new infections together with a recent rise in antiretroviral therapies (ARTs) halted the rise in the estimated number of AIDS deaths at about 2.2 million per year—equivalent to 4% of all global deaths (2). Among adults 15 to 49, the proportion currently infected with HIV (HIV prevalence) plateaued at just under 1% before declining to 0.8% worldwide (1, 3). These trends raise

the question of how global health funding should be rebalanced between AIDS treatment and HIV prevention, as well as other health-care investments.

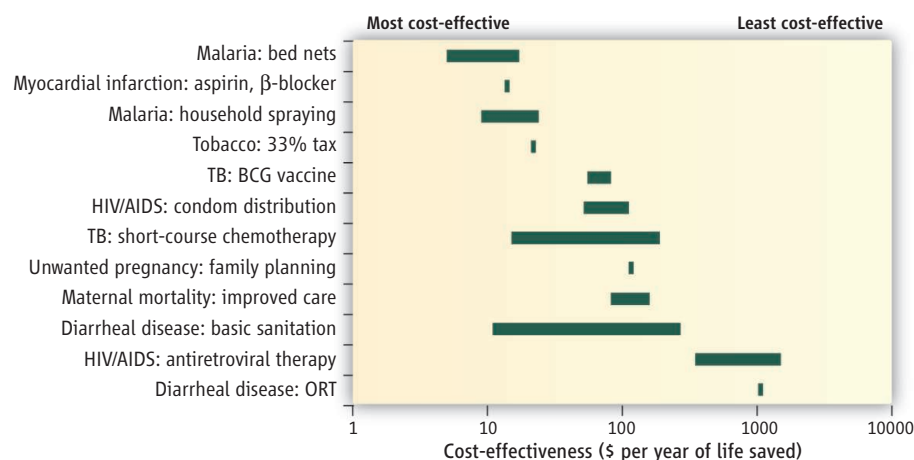
The cost of universal access to treatment is unsustainable. Medical and ethical considerations endow each patient currently on treatment with a life-long “entitlement” to receive at least his or her current treatment regimen (4, 5). Despite rapid growth in resources, less than half of those in need receive treatment, and five new infections occur for each two new persons put on treatment (3, 6). The World Health Organization (WHO) revised its recommendations regarding when to start treatment, raising the threshold from 200 CD4 cells/μl to 350, which could triple the number of people currently needing treatment (CD4 is a type of white blood cell that is killed during HIV

infection) (7). Reaching these ambitious targets would require the United States to spend half of its current foreign aid budget on AIDS treatment by 2016 and all of it by 2024 (4, 5).

The current allocation of health assistance to developing countries is far from optimal. One would expect resources allocated to a particular disease to be roughly proportional to the potential ill health averted by those expenditures. But the proportion of development assistance for health that is allocated to HIV/AIDS reached 23% in 2007, whereas the proportion of deaths attributable to AIDS in the developing world is less than 5% (3, 8). In a few African countries, foreign HIV/AIDS assistance exceeds the entire budget of the Ministry of Health (9). The huge influx of donor funding for HIV/AIDS sometimes crowds out other

¹Population Council, New York, NY 10017, USA. ²Center for Global Development, Washington, DC 20036, USA.

*Author for correspondence. E-mail: jbongaarts@popcouncil.org



Public Health Interventions. Cost-effectiveness estimates in low- and middle-income countries are illustrative of key interventions widely implemented and recommended by international organizations. For details on these and other interventions, see the supporting online material and (14). TB, tuberculosis; BCG, Bacille Calmette-Guérin vaccine for tuberculosis; ORT, oral rehydration treatment. Life-years saved are disability adjusted.

health needs and distorts health priorities, in part, by putting pressure on a meager supply of doctors, nurses, and clinics (4, 9–13).

To maximize a population's health status for a given amount of funding, the international donor community is ethically obliged to spend foreign aid funds and allocate health-care resources as cost-effectively as possible (13). Selected estimates of cost-effectiveness of interventions from the Disease Control Priorities Project (14) are shown in the chart, above. The range is very wide; ARTs for AIDS is one of the least cost-effective. Annually, 15.9 million people die from communicable diseases other than HIV/AIDS and from maternal and perinatal conditions and nutritional deficiencies, most preventable at low cost (15).

Even though institutional change is slow, a review of the U.S. President's Emergency Plan for AIDS Relief (PEPFAR) recommended a shift from "emergency relief" to "capacity building ... for sustainability" and to "greater emphasis on prevention of HIV" (16). The Obama Administration announced in December 2009 that, in addition to funding more treatment, it will accept this recommendation because "preventing new infections represents the only long-term, sustainable way to turn the tide against HIV/AIDS" (17). In January 2010, U.S. Secretary of State Clinton affirmed this rebalancing of health priorities by aiming to "invest \$63 billion over the next six years to help our partners improve their health systems and provide the care their people need rather than rely on donors to keep a fraction of their population healthy while the rest go with hardly any care" (18). This is a sig-

nificant shift from the previous administration's approach.

A World Bank evaluation of its programs warns that the dramatic increase in funds earmarked for HIV/AIDS "may be creating distortions in the rest of the health care system" and concludes that "performance of the HIV/AIDS portfolio has been much lower than that of other HPN [health, population, and nutrition] projects" (19). The overall objective of international donors should be transition to an HIV/AIDS policy that preserves recently achieved mortality reductions while lowering the annual number of new infections to less than the annual number of AIDS deaths. Sustainable progress can be made by donors specifying commitments to support existing ART patients, as well as a percentage (reduced from current levels) of those newly needing treatment. This will allow all stakeholders to plan for the future and for countries to decide how much additional treatment to finance themselves. In addition, donors could maintain the current treatment threshold of 200 CD4 cells/ μ l; integrate treatment within existing health systems; use incentives (e.g., cash or food) to raise adherence to treatment; and aggressively cut the cost of providing ARTs and of administration and technical assistance from outside the developing world.

Prevention efforts must be boosted to more than compensate for any increased transmission due to scaling back of the growth of ART access, in part by improving the monitoring of epidemic trends, rewarding countries for improvements (19, 20), and making evidence of effective prevention a condition for continuing support of

treatment. A range of prevention strategies is available (1, 20, 21). A recent review concludes that efforts to circumcise males and reduce the number of people who have sex with multiple partners appear to be particularly effective (21). Finally, donors must protect and expand resources for the most cost-effective health interventions, focusing on HIV prevention, childhood immunization, malaria, tuberculosis, maternal mortality, and family planning. These efforts will improve global health for a few dollars per year of life saved, instead of postponing deaths at hundreds of dollars per year saved with ARTs.

References and Notes

- UNAIDS, "AIDS Epidemic Update: December 2007" (UNAIDS, Geneva, 2007).
- J. Bongaarts, F. Pelletier, P. Gerland, *Lancet* **375**, 103 (2010).
- UNAIDS, "AIDS Epidemic Update: November 2009" (UNAIDS, Geneva, 2009).
- M. Over, *Rev. Econ. Dev.* **23**, 107 (2009).
- An English version of (4) is available at www.cgdev.org/content/publications/detail/15973.
- WHO, UNICEF, UNAIDS, *Towards Universal Access* (WHO, Geneva, 2009).
- WHO, *Rapid Advice: Antiretroviral Therapy for HIV Infection in Adults and Adolescents* (WHO, Geneva, 2009); www.who.int/hiv/pub/arv/rapid_advice_art.pdf.
- N. Ravishanker et al., *Lancet* **373**, 2113 (2009).
- J. Shiffman, *Health Policy Plan.* **23**, 95 (2008).
- M. Over, in *The Macroeconomics of HIV/AIDS*, M. Haacker, Ed. (International Monetary Fund, Washington, DC, 2004), pp. 311–344.
- M. Over, "The global AIDS transition: A feasible objective for AIDS policy" (Center for Global Development, Washington, DC, 2010); www.cgdev.org/content/publications/detail/1424143/.
- R. England, *BMJ* **336**, 1072 (2008).
- C. C. Denny, E. J. Emanuel, *JAMA* **300**, 2048 (2008).
- D. T. Jamison et al., *Disease Priorities in Developing Countries* (Oxford Univ. Press, Oxford, 2006).
- WHO, *The Global Burden of Disease: 2004 Update* (Who, Geneva, 2008); www.who.int/healthinfo/global_burden_disease/2004_report_update/en/index.html.
- Institute of Medicine, *PEPFAR Implementation: Progress and Promise* (National Academies Press, Washington, DC, 2007).
- Office of the U.S. Global AIDS Coordinator, *The U.S. President's Emergency Plan for AIDS Relief: Five-Year Strategy Annex* (U.S. Global AIDS Coordinator and Department of State, Washington, DC, 2008); www.pepfar.gov/documents/organization/133434.pdf.
- H. R. Clinton, speech at the Global Development Center, 6 January 2010; www.cgdev.org/content/general/detail/1423520.
- Independent Evaluation Group of the World Bank, *Improving Effectiveness and Outcomes for the Poor in Health Nutrition and Population* (World Bank, Washington, DC, 2009).
- M. Over, "Using Incentives to Prevent HIV Infections" (Center for Global Development, Washington, DC, 2010); www.cgdev.org/content/publications/detail/1424161/.
- M. Potts et al., *Science* **320**, 749 (2008).
- M.O. was supported by a grant to the Center for Global Development from the Bill and Melinda Gates Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5983/1359/DC1

10.1126/science.1191804

PALEONTOLOGY

Warm-Blooded “Sea Dragons”?

Ryosuke Motani

When dinosaurs roamed the land in the Mesozoic (251 to 65 million years ago), the top predators in the ocean were reptiles (1, 2). Three lineages of Mesozoic marine reptiles (plesiosaurs, ichthyosaurs, and mosasaurs) were especially successful (2) (see the first figure). They were similar to current marine mammals in many respects. They fed on fish, cephalopods, bivalves, and other air-breathing vertebrates (1). Ichthyosaurs evolved dolphin-like body plans. Plesiosaurs became underwater fliers, vaguely resembling sea lions (2, 3). It now appears that similarities to today's marine mammals extended further: On page 1379 of this issue, Bernard *et al.* (4) report that some ancient reptiles may have been able to sustain a constant body temperature (i.e., homeothermy).

Higher metabolic rates for these ancient reptiles, relative to modern ones, have previously been suggested, on the basis of bone

histology (5) and swimming energetics (6). Using oxygen isotopic data ($\delta^{18}\text{O}$), Bernard *et al.* now suggest that ichthyosaurs and plesiosaurs could sustain a body temperature (T_b) of up to $35^\circ \pm 2^\circ\text{C}$, whereas the T_b of mosasaurs changed with ambient temperature (T_a)—a thermal status called poikilothermy. For ichthyosaurs and plesiosaurs, this agrees with the idea that they were cruisers whose homeothermy allowed them to maintain constant muscle activity most efficiently. By contrast, homeothermy was not as useful in mosasaurs, which were ambush predators (2, 3, 6).

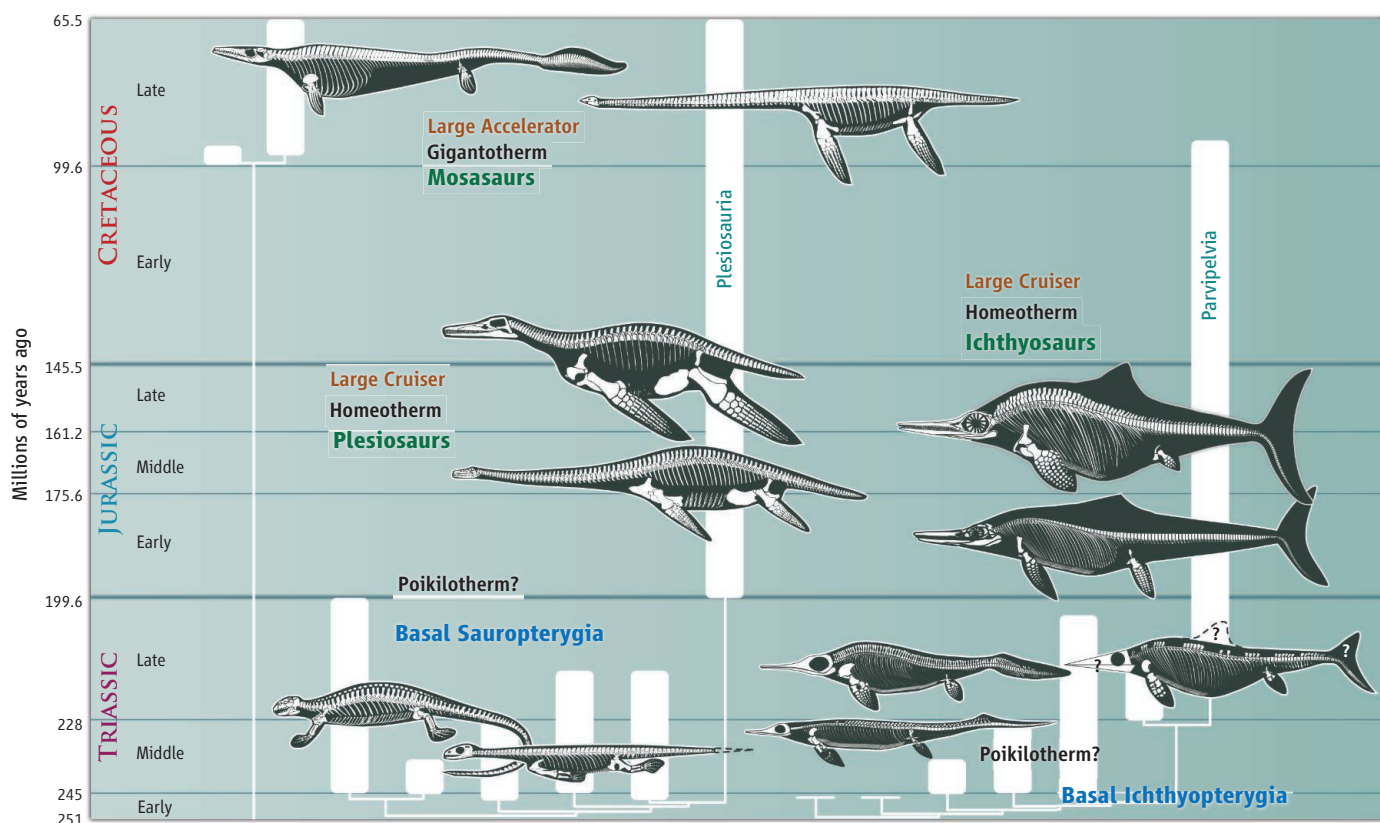
To estimate the temperature of ancient marine reptiles, the authors created a “thermometer” based on equations relating $\delta^{18}\text{O}$ and T_b in living fish. The equation is only applicable when a marine reptile had a T_b that equaled fish T_b ($= T_a$)—that is, it last grew in the water that was as warm as its T_b . Given that such a “lucky preservation” is rare, multiple $\delta^{18}\text{O}$ samples of both ancient fish and reptiles representing different time periods and places were analyzed to extrapolate this special T_b

Marine reptiles from the Age of Dinosaurs could maintain a constant body temperature.

value. This “fish thermometer” suggested that some large poikilothermic fish were as warm as $36^\circ \pm 2^\circ\text{C}$ (4). Even though some fish might survive in waters that warm, they would not be able to grow under such stressful conditions. Using this temperature scale, Bernard *et al.* argued that average mosasaurs had a T_b of $39^\circ \pm 2^\circ\text{C}$. If so, the T_b of an average mosasaur would rise with T_a until it equaled the latter at 39°C , implying that mosasaurs had trouble removing excess body heat.

Caution is required, as these temperatures are probably artifacts arising from time-dependent depletion of $\delta^{18}\text{O}$. Veizer *et al.* (7) noticed a linear decrease in $\delta^{18}\text{O}$ of about 0.02‰ per million years in invertebrate biogenic carbonates, possibly caused by tectonic activities (see the second figure). This means that the older the fossil, the higher the estimated temperature unless the bias is removed.

Three lineages of Mesozoic marine reptiles. White boxes indicate temporal extents of major subgroups, and connecting lines depict evolutionary relationships.



CREDIT: ADAPTED FROM R. MOTANI

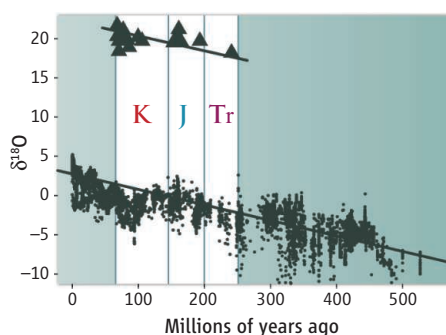
The long-term depletion trend is obscured in the Mesozoic part of the data (white background in the second figure) because it was unusually warm (i.e., lower $\delta^{18}\text{O}$) in the Cretaceous. Age depletion is not peculiar to biogenic carbonates but is seen in vertebrate phosphates as well (see the second figure).

Compensating for this trend would lower the warmest fish growth temperature to $26^\circ \pm 2^\circ\text{C}$, and ichthyosaurian T_b to about $24^\circ \pm 2^\circ\text{C}$. The values agree with our knowledge of marine vertebrate physiology (excluding mammals and birds): The T_b of homeothermic marine vertebrates, such as tunas and leatherback turtles, is about 20° to 25°C (8, 9). A growth temperature of 26°C occurs among fish living in the warmest seas today.

The conclusion of homeothermy was not affected by this bias removal. Although future scrutiny is necessary, the following picture emerges. At a T_a of 10°C , all three groups had a T_b of about 24°C . As the T_a rose to 36°C , mosasaur T_b increased to 36°C , whereas the other two groups maintained an almost constant T_b . Large cruisers among living nonmammalian marine vertebrates red muscles tuned for an in vivo temperature centered around 25°C (9, 10).

All three groups had a higher T_b than co-occurring fish by about 5° to 20°C , with the exception of Triassic ichthyosaurs. This suggests that they had heat conservation systems such as blubber layers and specialized blood circulation. Perhaps some of them were endothermic (i.e., having elevated resting metabolic rates and generating extra heat inside their bodies). However, no living reptile (other than birds) is known to be endothermic. Even homeothermic leatherback turtles are not endothermic but gigantothermic [i.e., they maintain T_b by virtue of large body size and insulation], sometimes more than 20°C above T_a (8), without extra internal heat generation. The warmest marine reptiles described by Bernard *et al.* were giants, such as the mosasaur *Tylosaurus*, whereas a co-occurring smaller mosasaur was about 5°C colder. Considering this and the overheating tendency of mosasaurs, they may have been gigantothermic. Perhaps plesiosaurs and ichthyosaurs were regional endotherms, as are tunas and lamnid sharks (9).

The conclusions of Bernard *et al.* only apply to Jurassic and Cretaceous marine reptiles, given the data range. By that time, ichthyosaur and plesiosaur lineages had completed their adaptation for cruising, with ichthyosaurs having semi-crescent-shaped tail flukes and plesiosaurs two pairs of rigid and pointed flippers (see the first figure). When did homeothermy evolve in these lineages?



Similarity in $\delta^{18}\text{O}$ depletion trends between biogenic carbonate (triangles) and phosphates (dots). Lines represent standardized major axis regression lines as implemented in the smatr package of software platform R (13). Data were taken from (4) and (7). K, Cretaceous; J, Jurassic; Tr, Triassic.

They lived on land until the Early Triassic and, given the physiology of current land-living reptiles, it is unlikely that the land ancestors were homeothermic. The transition may have occurred in the Late Triassic, as these reptiles thrived through changing environments after their initial aquatic adaptation.

Evolution of a tuna-shaped body plan in ichthyosaurs, which enabled them to cruise the seas, was correlated with a decrease in coastal habitats in the Middle to Late Trias-

sic (11). Most coastal marine reptiles became extinct as their habitats were lost, but cruising ichthyosaurs prospered. If homeothermy in ichthyosaurs evolved in conjunction with cruising, then there is an opportunity to clarify environmental effects on evolution. Evolutionary triggers for homeothermy are generally difficult to decipher (12), yet comparisons of the evolutionary records of marine reptiles and fish may illuminate common environmental factors.

References and Notes

1. J. A. Massare, *J. Vertebr. Paleontol.* **7**, 121 (1987).
2. R. Motani, *Evol. Educ. Outreach* **2**, 224 (2009).
3. J. A. Massare, in *Mechanics and Physiology of Animal Swimming*, L. Maddock, Q. Bone, J. M. V. Rayner, Eds. (Cambridge Univ. Press, Cambridge, 1994), pp. 133–149.
4. A. Bernard *et al.*, *Science* **328**, 1379 (2010).
5. V. D. Buffrénil, J. Mazin, *Paleobiology* **16**, 435 (1990).
6. R. Motani, *Nature* **415**, 309 (2002).
7. J. Veizer, Y. Godderis, L. M. François, *Nature* **408**, 698 (2000); updated data available from www.science.uottawa.ca/~veizer/isotope_data/.
8. B. P. Wallace, T. T. Jones, *J. Exp. Mar. Biol.* **356**, 8 (2008).
9. J. B. Graham, K. A. Dickson, *J. Exp. Biol.* **207**, 4015 (2004).
10. D. Bernal *et al.*, *Nature* **437**, 1349 (2005).
11. R. Motani, *Sociol. Pap.* **14**, 147 (2008).
12. K. A. Dickson, J. B. Graham, *Physiol. Biochem. Zool.* **77**, 998 (2004).
13. D. I. Warton *et al.*, *Biol. Rev. Camb. Philos. Soc.* **81**, 259 (2006).

10.1126/science.1191409

APPLIED PHYSICS

Pulling Apart Molecular Magnetism

Pablo Jarillo-Herrero

Mechanical control can be used to probe and tune the magnetic properties of single molecules.

A single molecule constitutes the ultimate nanometer-scale object through which electronic transport can take place. Being so small, molecules share many characteristics with atoms, such as discrete quantized energy spectra and angular momentum, yet at the same time are large enough to be mechanically deformed and chemically attached to metallic leads with which they can exchange electrons. Can these mechanical and exchange interactions be controlled and, if so, what new phenomena arise? On page 1370 of this issue, Parks *et al.* (1) show how magnetism and quantum many-body phenomena can be tuned by precise mechanical manipulation of single molecules.

Quantum mechanics tells us that electrons confined within a potential “box” can have a discrete set of allowed energies. Such

a discrete energy spectrum (the energy levels and the spacing between them) depends on the strength of the confining potential and the boundary conditions imposed on the electron wave functions. For symmetric potentials, such as the Coulomb potential from a nucleus, the quantum energy states exhibit orbital degeneracies (i.e., the s, p, d... atomic shells). Pauli’s exclusion principle dictates that at most two electrons with total spin zero can occupy a given energy state. But for orbitally degenerate states, two or more electrons may align their spins by occupying different states, minimizing their Coulomb repulsion—as is exemplified by Hund’s rule. Orbital degeneracies are of fundamental importance for the existence of atoms with large magnetic moments (spin, $S \geq 1$) and play an important role in molecular magnetism.

Surrounding such magnetic atoms with other atoms and ligands to form a molecule alters the boundary conditions for the quan-

Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: pjarillo@mit.edu

tum box. This can lead to a splitting of the different orbital and spin states, depending on the symmetry properties of the atom's environment, and in turn results in a modification of the magnetic properties of the entire system. Parks *et al.* investigate the magnetic states of individual $S = 1$ molecules placed between two electrodes in a metal break junction. By stretching the electrodes, they explore the resulting symmetry-breaking effects of this mechanical action on the conduction properties through the molecule. They focus on the magnetic anisotropy of the molecules—how the energy differences among the different S_z states ($S_z = 0, \pm 1$) change for different degrees of stretching.

Now, getting to know the spin configuration of an individual molecule in a changing environment is nontrivial, but here is where quantum many-body physics comes to the rescue. A well-known correlated electronic trans-

is twofold degenerate (both $S_z = \pm 1/2$ spin orientations have the same energy). Creating an energy splitting between these spin states by, for example, applying a magnetic field, leads to a splitting of the Kondo resonance as soon as the splitting exceeds $\sim T_K$.

At zero temperature, a number $2S$ of channels in the leads are required to completely screen an isolated spin S (S). In the case of a spin- $1/2$ impurity, this means just one screening channel, which typically consists of combinations of conducting states in the two metallic leads to which the molecule is coupled. For larger spin, for example, $S = 1$, things get trickier. Each of the two screening channels ($2S = 2$) will have in general a different coupling to the magnetic impurity. Because the Kondo temperature associated with each screening channel is an exponential function of the coupling strength, this means that there will be two Kondo tempera-

called the magnetic anisotropy energy, D (see the figure, panel B). Because the stretching can be controlled with subangstrom precision, the anisotropy energy is exquisitely tunable. Once the energy splitting is comparable to T_K , the degeneracy is broken such that the Kondo effect ceases to occur. Parks *et al.* observe a gradual splitting of the Kondo conductance peak as the distance between the electrodes is varied, in agreement with prediction.

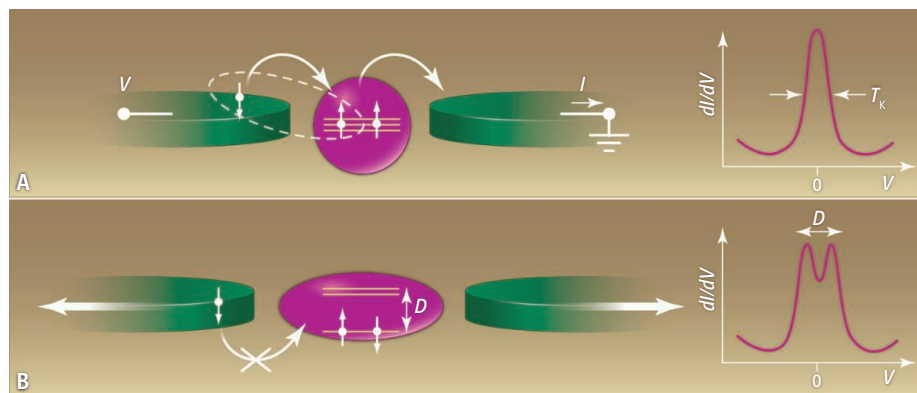
The Kondo effect has been used as a spectroscopic tool to study magnetic anisotropy in magnetic atoms connected to an environment (8). However, in these previous studies the atoms had a half integer spin ($S = 3/2$) for which time reversal symmetry guarantees partial spin degeneracy at zero magnetic field, and therefore a single, unsplit Kondo peak at zero magnetic field. The study by Parks *et al.* thus constitutes the case where a Kondo effect associated with a magnetic atom can be tuned by purely mechanical means, resulting in a new degree of control to investigate Kondo resonances and splittings at zero magnetic field. Moreover, having two independent knobs to tune the energy spectrum—mechanical pulling for the anisotropy and magnetic field for the Zeeman splitting—they confirm the theoretically expected behavior for the Kondo resonance peak splitting under simultaneous magnetic field and mechanical stretching.

The results by Parks *et al.* show that the Kondo effect, one of the most researched quantum many-body phenomena in strongly correlated physics, continues to surprise in all its variations (spin, orbital, multichannel, and here underscreened), and now too as a superb spectroscopic tool for nontrivial interrogations of the magnetic behavior of individual molecules. The ability to mechanically tune the magnetic anisotropy of individual molecules provides the opportunity to systematically investigate and theoretically compare quantum chemistry models, which will enormously broaden our knowledge of molecular magnetism. Beyond the fundamental science level, it may play an important role in the use and manipulation of nanoscale magnets for spintronics and quantum information processing applications.

References and Notes

1. J. J. Parks *et al.*, *Science* **328**, 1370 (2010).
2. L. Kouwenhoven, L. Glazman, *Phys. World* **14**, 33 (2001).
3. D. Goldhaber-Gordon *et al.*, *Nature* **391**, 156 (1998).
4. S. M. Cronenwett *et al.*, *Science* **281**, 540 (1998).
5. P. Nozieres, A. Blandin, *J. Phys.* **41**, 193 (1980).
6. A. Posazhennikova, P. Coleman, *Phys. Rev. Lett.* **94**, 036802 (2005).
7. N. Roch *et al.*, *Phys. Rev. Lett.* **103**, 197202 (2009).
8. A. F. Otte *et al.*, *Nat. Phys.* **4**, 847 (2008).
9. I acknowledge useful discussions with R. C. Ashoori and D. Goldhaber-Gordon.

10.1126/science.1191411



Magnetism on the pull. (A) For the unstretched molecule, the three triplet states are degenerate, leading to a Kondo peak in dI/dV at $V = 0$. (B) Mechanically stretching the molecule splits the triplet states' degeneracy, thus leading to a controllable split of the Kondo peak, as studied by Parks *et al.*

port phenomenon, called the Kondo effect (2), is exquisitely sensitive to the presence of spin and orbital degeneracies. In its simplest incarnation—the spin $1/2$ Kondo effect—an isolated spin- $1/2$ magnetic impurity (e.g., a single electron) is antiferromagnetically screened by the conduction electrons in a host metal. The net result is the formation of a singlet state (a spin-up and spin-down combination, with $S = 0$) between the metallic leads and the magnetic impurity. The Kondo effect results in a continuous exchange of spin-up and spin-down electrons between the leads and the magnetic impurity (3, 4). This exchange enables highly efficient electronic transport through the impurity at zero bias, that is, a single peak in differential conductance (dI/dV) at a source-drain bias (V) of zero (see the figure, panel A). The binding energy of this correlated state is characterized by a temperature scale called the Kondo temperature, T_K . Key to this process is that the spin- $1/2$ impurity state

tures, T_{K1} and T_{K2} , which can be very different. Therefore, in general there will be a temperature regime $T_{K1} < T < T_{K2}$, in which only one out of the two spin- $1/2$ degrees of freedom in the molecule will be screened. This results in the underscreened Kondo effect, which has a very different temperature dependence than the ordinary $S = 1/2$ Kondo effect (6). The strategy then is to measure the differential conductance of your molecular device. If you see a Kondo peak, then its temperature dependence reveals the spin of the molecule.

The theoretical predictions for this $S = 1$ underscreened Kondo effect were recently confirmed in a different experiment (7). Parks *et al.* report that they can lift the degeneracy (separate the levels) of the three triplet states by mechanically pulling apart the metal leads connected to the molecule. Such molecular stretching results, in this case, in a $S_z = 0$ singlet ground state, with the two other triplets ($S_z = \pm 1$) being higher in energy by an amount

Think Vesicular Chloride

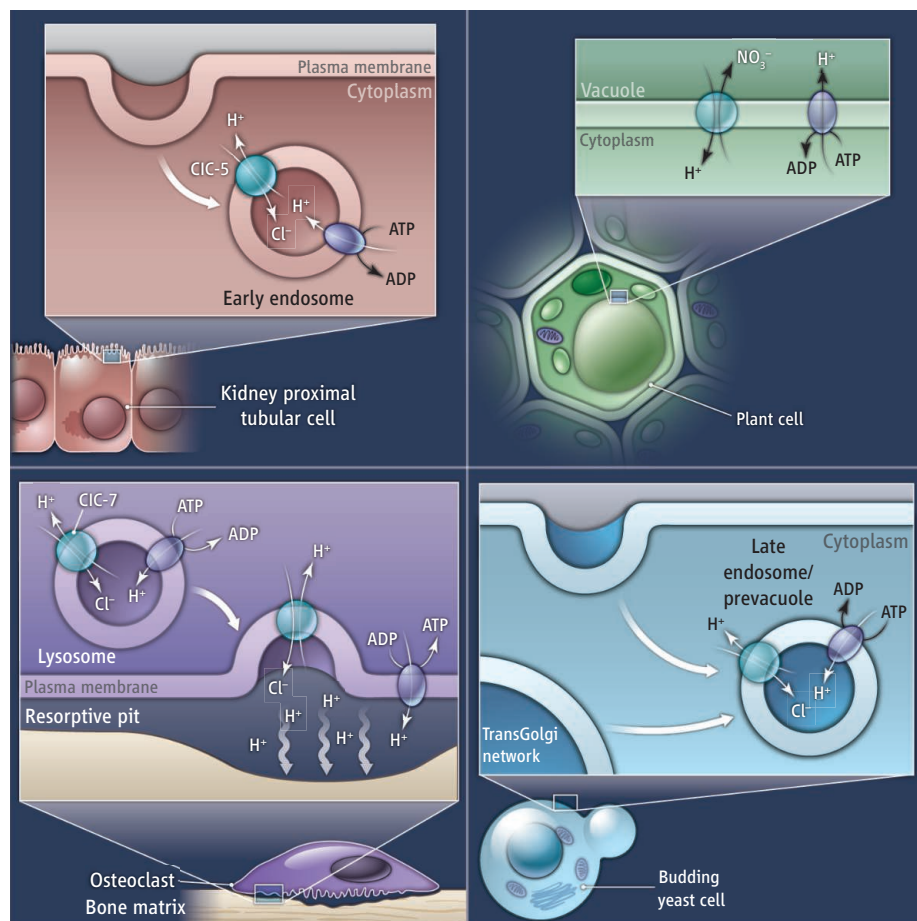
Andrew J. Smith and Blanche Schwappach

The distinction between passive and active ion movement through cellular channels and exchangers, respectively, has rarely been more relevant to understanding the physiological function of a protein than in the case of intracellular members of the chloride channel (CLC) family. Different CLC family members act either as chloride ion channels or as chloride-proton exchangers. On pages 1398 and 1401 of this issue, Novarino *et al.* (1) and Weinert *et al.* (2) demonstrate the relevance of the exchange activity in mammalian cells.

According to the prevalent hypothesis for the role of CLC proteins in endocytosis in kidney tubular cells (CLC-5) and in lysosomal biology (CLC-7), chloride moves into the organelle lumen via the CLC protein toward a positive electrical potential that a primary active proton pump has generated across the membrane. This movement of chloride dissipates the potential and enables vesicle acidification (“chloride shunt hypothesis”). Instead, Novarino *et al.* and Weinert *et al.* hypothesized that the transporter exploits the proton gradient to concentrate chloride inside lysosomes and endosomes (see the figure). Such activity was shown for the CLC-mediated accumulation of nitrate in the plant vacuole (3).

Mice genetically engineered to lack CLC-5 (*Cln5^{-/-}*) or CLC-7 (*Cln7^{-/-}*) are of interest because their pathology mirrors the symptoms of human disorders. The *Cln5^{-/-}* mouse is a good model of the kidney symptoms occurring in Dent's disease, and *Cln7^{-/-}* mice develop osteopetrosis (abnormally shaped bones) and neuronal degeneration as observed in human patients (4). Novarino *et al.* and Weinert *et al.* compared these mice with mice engineered to express an altered form of either CLC in which a highly conserved glutamate is mutated (*Cln5^{unc/unc}* and *Cln7^{unc/unc}*). When this residue is altered, proton and chloride exchange becomes uncoupled, and both CLCs function as passive chloride channels (5). This approach allowed the authors to determine the relevance of the exchange activity to the pathology of either mouse. In both cases, the main symptoms

Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK. E-mail: blanche.schwappach@manchester.ac.uk



Coupled exchange. Intracellular members of the CLC family are critical to the function of subcellular organelles in mammalian, plant, and yeast cells.

of the mice lacking either CLC protein were generally recapitulated in mice expressing the uncoupled form of either exchanger. The result thus excludes the chloride shunt hypothesis. Osteopetrosis was milder in the *Cln7^{unc/unc}* mice than in wild-type animals, indicating that although the uncoupled transporter only functions as a chloride channel, it still serves some residual function.

The importance of coupled transport by an intracellular CLC is also observed in baker's yeast. The single yeast CLC protein, Gef1, is required for the maturation of multicopper oxidases in a late endosomal/prevacuolar compartment. When its chloride-proton exchange activity is uncoupled (in a mutant *gef1^{unc}* strain), the mutant protein presumably acts as a passive chloride channel and the oxidases do not mature properly (6). Failure of the

Transporters that import chloride ions in exchange for the export of protons control the function of intracellular vesicles in mammalian cells.

uncoupled CLC variants to replace the wild-type proteins in mice shows their importance to endosome and lysosome function, but new issues are raised by their residual function. As these variants likely retain all properties of the wild-type proteins except for exchange activity, any differences between wild-type and *unc/unc* mice point toward roles that a chloride channel can fulfill—either by allowing passive chloride movement or by providing an interaction site for other proteins that contribute to organelle function.

It was previously proposed that CLC-5 could directly acidify early endosomes by exploiting higher concentrations of chloride inside the compartment. This gradient of chloride would drive CLC-5 to import protons—so-called secondary active acidification (5). Indeed, recent work demonstrates that CLC-5

controls endosomal acidification in a manner that is independent of a separate proton pump (7). It is unclear whether this or lack of vesicular chloride explains the endocytic defects of *Clcn5^{unc/unc}* mice. With the work by Weinert *et al.*, the field of organellar physiology has, however, taken an exciting turn. The authors report decreased chloride concentration inside *Clcn7^{unc/unc}* lysosomes but normal lysosomal pH. This implicates luminal chloride as an important parameter in endolysosomal physicochemistry. Current methodology to assess ion gradients and membrane potential across intracellular membranes *in vivo* is severely limited, and improvements in live cell imaging of these parameters will have a major impact on the field.

The findings of Weinert *et al.* support a physiological requirement for chloride along the endosomal and lysosomal pathway and raise an important question in the field: What is the function of vesicular chloride? Is it required as a cofactor of protein biogenesis as has been suggested for the yeast model (8)? Or does it control the concentration of other physiologically relevant anions by influencing their ability to move across the relevant membrane (6)? Distinct chemical reaction spaces are generally considered one major advantage of maintaining subcellular organelles. The life of organelles is not static, and the processes controlled by intracellular ion channels and exchangers may be subject to regulation, such as by metabolism (6, 9, 10).

References

1. G. Novarino, S. Weinert, G. Rickheit, T. J. Jentsch, *Science* **328**, 1398 (2010); published online 29 April 2010 (10.1126/science.1188070).
2. S. Weinert *et al.*, *Science* **328**, 1401 (2010); published online 29 April 2010 (10.1126/science.1188072).
3. A. De Angeli *et al.*, *Nature* **442**, 939 (2006).
4. R. Planells-Cases, T. J. Jentsch, *Biochim. Biophys. Acta* **1792**, 173 (2009).
5. C. Miller, *Nature* **440**, 484 (2006).
6. N. A. Braun, B. Morgan, T. P. Dick, B. Schwappach, *J. Cell Sci.* **10.1242/jcs.06843** (2010).
7. A. J. Smith, J. D. Lippiat, *J. Physiol.* **10.1113/jphysiol.2010.188540** (2010).
8. S. R. Davis-Kaplan, C. C. Askwith, A. C. Bengtzen, D. Radisky, J. Kaplan, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13641 (1998).
9. S. Meyer, S. Savaresi, I. C. Forster, R. Dutzler, *Nat. Struct. Mol. Biol.* **14**, 60 (2007).
10. A. De Angeli *et al.*, *J. Biol. Chem.* **284**, 26526 (2009).

10.1126/science.1191529

CHEMISTRY

Clean, Green Chiral Reactions—Just Add a Salt

Andrew N. French

Chirality—literally handedness—refers to the mirror-image, or left-right asymmetry of objects like shoes and gloves. Like our hands, many chemical compounds (and most biomolecules) exist as enantiomers, which, relative to each other, have mirror-image arrangements of substituents around an atom (typically carbon or nitrogen). In many cases, pharmaceuticals must have a specific chirality in order to be active, but making pure chiral compounds is difficult and costly. Traditionally, chemists have relied on reagents for controlling chirality that contain heavy metals (1), the presence of which can lead to toxicity issues. A relatively new alternative is the use of nonmetal catalysts called organocatalysts. On page 1376 of this issue, Uyanik *et al.* (2) report a more “green” organocatalytic reaction that also uses an iodine reagent to close rings in a molecule during an oxidation step.

Reactions that produce chiral products—called asymmetric reactions—are evaluated in terms of their selectivity. The metric used is the “enantiomeric excess,” or ee, which is the excess of one form of the product over the other. Despite the environmental drawbacks of using metals such as osmium, chromium, or titanium, the selectivity of metal-containing catalysts approaches 98% or greater ee

for some transformations. Alternatives that have much lower ee’s will not be “green” if they are inefficient and waste reagents.

Iodine, like many metals, has the ability to become hypervalent; it can expand the number of valence electrons beyond the usual octet. Hypervalent iodine reagents have been used in many chemical reactions, including oxidations and functionalizations of multiple bonds, aromatic compounds, and a host of other substrates (3, 4). Chiral hypervalent iodine reagents have been an active area of study since the 1980’s, whereas the realm of aryl iodides as organocatalysts is a relatively new area of research (5–11). These organic compounds share a structural feature; the iodine atom doing the catalysis is attached to an aromatic ring, with the chirality incorporated onto that ring system in some way. Uyanik *et al.* have, in fact, recently explored the use of chiral aryl iodides as organocatalysts (12).

In the present study, Uyanik *et al.* took a different approach to hypervalent iodine chemistry. The oxidative power of hydrogen peroxide is exploited in that I^- is converted in aqueous solution to either hypoiodite (IO^-) or iodite (IO_2^-) (see the figure, panel A). To use these species in asymmetric organic reactions, they need to be chaperoned into organic solvents, and they need help in imparting chirality. Uyanik *et al.* chose a chiral quaternary ammonium cation to solve these problems

An ammonium iodide salt can perform selective organic oxidation reactions and replace more toxic metal catalysts.

(13). Upon binding the hypoiodite (IO^-) or iodite (IO_2^-), a neutral complex forms that can dissolve in both the aqueous and the organic phases. The shape and chiral nature of the quaternary ammonium cation (see the figure, panel B) is the key to imparting chirality in the product. Such molecules can interact with the organic substrate so that the oxidation reaction favors one of the handed products over the other.

Uyanik *et al.* reacted ketophenols, compounds with OH-substituted benzenes that also have a C=O bond tethered nearby, with this catalyst and hydrogen peroxide. They performed this reaction on a number of substituted phenols and obtained very good yields and selectivities regardless of the pattern of substitution. The selectivities observed were in excess of 85% ee, and greater than 90% for monosubstituted substrates (some examples were in excess of 95% ee). These selectivities represent the highest reported for a hypervalent iodine reagent-catalyzed reaction and are competitive with metal-catalyzed reactions in both yield and selectivity. Additionally, because the co-oxidant was hydrogen peroxide, the by-product from this oxidation was water.

Not only are the selectivities exciting, but the system is also much greener than metal-catalyzed reactions. The catalyst precursor is I^- , iodide, the same anion found in iodized salt. The co-oxidant is hydrogen peroxide,

Chemistry Department, Albion College, Albion, MI 49224, USA. E-mail: afrench@albion.edu

whose by-product after reaction is water.

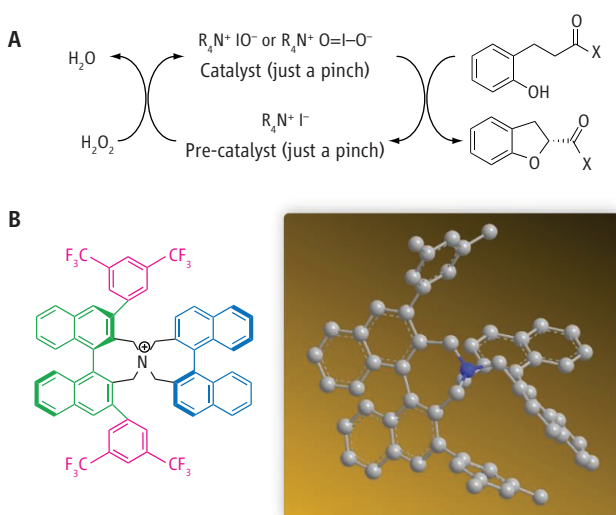
The products of the catalytic reaction described by Uyanik *et al.* create the framework for a number of natural products with varied biological effects. The benzofuran products can form the starting point for the synthesis of more complex pharmaceutical candidates. For example, tremetone has both antifungal and insecticidal properties and is derived from a plant extract. A similar plant

isolate, rotenone, is a potent anti-leukemic drug candidate as well (14). In 2005, the Merck Company reported the synthesis of a drug candidate with this same backbone that modulates the levels of serum triglycerides and high-density lipoprotein in the blood (15). A synthesis of this target with this new method could be accomplished in fewer steps than the 2005 method, produce less waste, and reduce cost.

The future of hypervalent iodine is likely to be as varied as the chemists working in this area. Elucidating the mechanism, including the steps that lead to a chiral product, will allow for further improvements in selectivity. Ideally, this catalyst system, like any catalyst developed, will be tested on other substrates and reactions involving iodine-containing catalysts. Iodine chemistry, with its versatile reactivity, is an excellent area to discover new, more environmentally friendly, greener organocatalysts. The chemistry described by Uyanik *et al.* is but a taste of what is to come.

Guiding iodine catalysts to their targets.

(A) In the reaction described by Uyanik *et al.*, hydrogen peroxide reacts with the salt formed by a chiral ammonium cation (R_4N^+) and iodide, making water and oxidized iodine. This hypervalent (hypo)iodite then reacts with the ketophenol to generate the chiral benzofuran skeleton, where X can be one of many different functional groups. The reaction shows excellent selectivity for one of the two products that differ in handedness (in this case, the favored product has the COX group behind the plane formed by the rest of the molecule; the other enantiomer has this group in front). The compound is a starting point for a host of pharmaceutical candidates. (B) The structural formula of the chiral ammonium cation is shown on the left. The three-dimensional rendering of the chiral ammonium salt on the right has the nitrogen atom in blue and carbon atoms in gray; hydrogen and fluorine atoms are omitted for clarity.



References

1. T. Katsuki, K. B. Sharpless, *J. Am. Chem. Soc.* **102**, 5974 (1980).
2. M. Uyanik, H. Okamoto, T. Yasui, K. Ishihara, *Science* **328**, 1376 (2010).
3. T. Wirth, *Angew. Chem. Int. Ed.* **44**, 3656 (2005).
4. R. M. Moriarty, *J. Org. Chem.* **70**, 2893 (2005).
5. T. Dohi *et al.*, *Chem. Commun. (Camb.)* **2005**, 2205 (2005).
6. T. Dohi *et al.*, *Angew. Chem. Int. Ed.* **44**, 6193 (2005).
7. C. I. Herreras, T. Y. Zhang, C.-J. Li, *Tetrahedron Lett.* **47**, 13 (2006).
8. Y. Yamamoto, H. Togo, *Synlett* **2006**, 798 (2006).
9. R. D. Richardson, T. Wirth, *Angew. Chem. Int. Ed.* **45**, 4402 (2006).
10. R. D. Richardson *et al.*, *Synlett* **2007**, 0538 (2007).
11. T. Dohi *et al.*, *Angew. Chem. Int. Ed.* **47**, 3787 (2008).
12. M. Uyanik, T. Yasui, K. Ishihara, *Angew. Chem. Int. Ed.* **49**, 2175 (2010).
13. K. Maruoka, *Org. Process Res. Dev.* **12**, 679 (2008).
14. M. Abou-Shoer, F. E. Boettner, C.-J. Chang, J. M. Cassidy, *Phytochemistry* **27**, 2795 (1988).
15. G. Q. Shi *et al.*, *J. Med. Chem.* **48**, 5589 (2005).

10.1126/science.1191408

ATMOSPHERIC SCIENCE

Getting to the Critical Nucleus of Aerosol Formation

Renyi Zhang^{1,2,3}

Atmospheric aerosols—microscopic particles suspended in Earth's atmosphere—are a major environmental problem. They degrade visibility, negatively affect human health, and directly and indirectly influence climate by absorbing and reflecting solar radiation and modifying cloud formation. Researchers do not, how-

ever, fully understand at the molecular level how aerosols form, creating one of the largest sources of uncertainty in atmospheric models and climate predictions (1). Recent findings suggest a path to a better understanding of aerosol formation (2–4).

Aerosols can be directly emitted into the atmosphere—for example by plants, combustion, or sea spray—or form through a chemical process known as nucleation, in which gaseous molecules bond. Nucleation produces a large fraction of atmospheric aerosols, and investigators have frequently observed nucleation in various environments, including urban, forested, and marine areas (5). New particle formation is commonly

A better understanding of how aerosols form in the atmosphere could greatly improve climate models.

considered to be a two-step process: First, nucleation forms a “critical nucleus,” which then grows to a detectable size (6). Classical nucleation theory reveals that when the critical nucleus forms, the free energy of the nucleating system reaches a maximum—“the nucleation barrier”—beyond which aerosol growth becomes spontaneous. The rate at which nucleation occurs is related to the chemical makeup of the critical nucleus and the gaseous concentrations of the nucleating species (7). That rate is an important variable in simulations of aerosol formation in atmospheric models (8).

Previous studies, however, have not been able to directly measure nucleation rates

¹State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University, Beijing, 100871, China. ²College of Environmental Science and Engineering, Fudan University, Shanghai, 200433, China. ³Departments of Atmospheric Sciences and Chemistry, Center for Atmospheric Chemistry and the Environment, Texas A&M University, College Station, TX 77843, USA. E-mail: renyi-zhang@tamu.edu

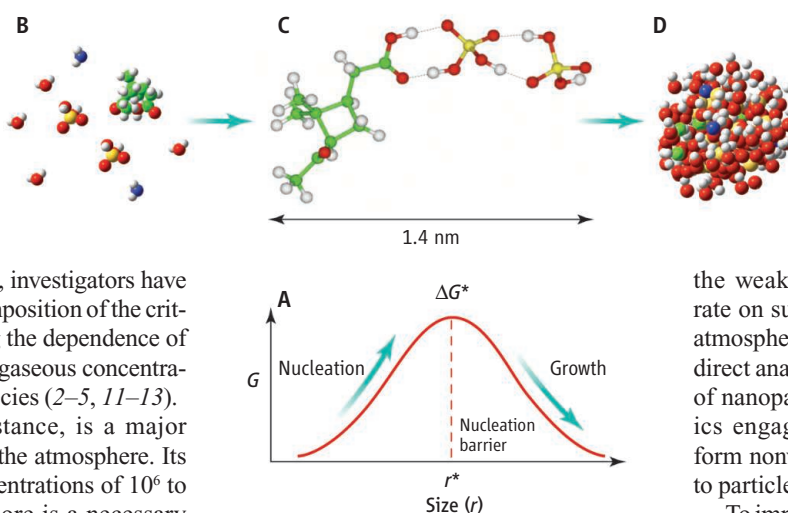
or the chemical composition of the critical nucleus in binary or multicomponent systems like those found in the atmosphere. Theoretical methods have also failed to reliably identify the nucleation

barrier (9, 10). As a result, investigators have indirectly inferred the composition of the critical nucleus by measuring the dependence of the nucleation rate on the gaseous concentrations of the nucleating species (2–5, 11–13).

Sulfuric acid, for instance, is a major nucleating component in the atmosphere. Its presence in gaseous concentrations of 10^6 to 10^7 molecules cm^{-3} or more is a necessary condition for new particle formation (5). Atmospheric measurements have suggested that nucleation rates depend weakly on sulfuric acid concentrations, implying that just one or two sulfuric acid molecules are present in the critical nucleus (5). In contrast, laboratory studies suggest that nucleation rates depend more strongly on sulfuric acid concentrations, corresponding to a critical nucleus of four to nine sulfuric acid molecules (11–13). This larger number agrees with predictions under classical nucleation theory (7).

In recent laboratory experiments, Sipilä *et al.* reported rapid binary nucleation of sulfuric acid at concentrations comparable to those found in the atmosphere; their finding implicated a critical nucleus consisting of one or two sulfuric acid molecules (2). The difference between these results and previous laboratory measurements is explained by the authors' use of an improved instrument that can count particles as small as 1.5 nm. Previous measurements were limited by a low counting efficiency for particles smaller than 3 nm, resulting in appreciable underestimates of the nucleation rate. Sipilä *et al.*'s conclusion that nucleation weakly depends on the concentration of sulfuric acid raises an important question: Are one or two sulfuric acid molecules (a monomer or dimer) enough to form a critical nucleus?

Several lines of evidence suggest that the answer should be “no.” Molecular dynamics simulation, for instance, suggests that a hydrated sulfuric acid dimer has a diameter of 0.7 nm, which is commonly believed to be too small to overcome the nucleation barrier (7). Quantum chemical calculations show the existence of two medium-strength hydrogen bonds in the sulfuric acid dimer, and the available thermodynamic data predict that the dimers would rapidly decompose under typical atmospheric concentra-



Organics-assisted aerosols. In aerosol formation, bonding particles must cross an energy threshold—the nucleation barrier—beyond which aerosol growth becomes spontaneous (A). Organic acids (B) (carbon in green) that mingle with gaseous sulfuric acid molecules (C) (sulfur in yellow) could facilitate the crossing of the barrier by creating a critical nucleus with one or two sulfuric acid molecules (C), leading to aerosol growth (D). Knowing the composition of the critical nucleus would enable researchers to predict the nucleation rate, an important variable in atmospheric models (16).

tions of sulfuric acid (10^6 to 10^8 molecules cm^{-3}) (9, 10).

Metzgera *et al.* and others, however, suggest that it is highly plausible that the answer is “yes” if other sulfuric acid-stabilizing species are involved in nucleation and are present in the critical nucleus (3, 8). One candidate is organic acids, because they form larger and more stable heterodimers with sulfuric acid (4, 9, 10, 12). The interaction between an organic acid and sulfuric acid involves one strong and one medium-strength hydrogen bond, with a binding energy that is 2 to 3 kcal mol^{-1} larger than that of the sulfuric acid dimer (9, 10), and the heterodimer has a vacant OH group in the sulfuric acid moiety to allow further growth through hydrogen-bond formation. A dimer of two organic acids also has a large binding energy (9), but no hydrogen acceptor or donor group is available for subsequent growth, so organic acid dimers contribute negligibly to new particle formation (4).

Several studies indicate that the critical nucleus consists of only one molecule of the organic species (3, 4). The presence of organic acids in laboratory-produced nanoparticles has been confirmed in particles as small as 4 nm (4). In addition, atmospheric concentrations of organic acids are expected to be much higher than that of sulfuric acid,

because of photochemical oxidation of volatile organic compounds abundantly emitted by biogenic and anthropogenic sources (8, 14). The presence of organic acids could enable fewer sulfuric acid molecules to form a critical nucleus (see the figure), and would explain

the weaker dependence of the nucleation rate on sulfuric acid concentration found in atmospheric measurements (3, 8). Moreover, direct analyses of the chemical compositions of nanoparticles have suggested that organics engage in heterogeneous reactions to form nonvolatile compounds that contribute to particle growth (15).

To improve models used to assess the environmental and climate impacts of aerosols, it is imperative for future studies to precisely quantify the chemical makeup of the critical nucleus. This may be accomplished by augmenting advanced theoretical approaches (i.e., quantum chemical or molecular dynamics simulations) with simultaneous measurements of the size and chemical composition of freshly nucleated nanoparticles in the laboratory and in the field.

References and Notes

1. IPCC, *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, UK), www.ipcc.ch/ipccreports/ar4-wg1.htm (2007).
2. M. Sipilä *et al.*, *Science* **327**, 1243 (2010).
3. A. Metzgera *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6646 (2010).
4. R. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17650 (2009).
5. P. H. McMurry *et al.*, *J. Geophys. Res.* **110**, D22S02 (2005).
6. M. Kulmala, L. Pirjola, J. M. Makela, *Nature* **404**, 66 (2000).
7. R. McGraw, R. Zhang, *J. Chem. Phys.* **128**, 064508 (2008).
8. J. Fan, R. Zhang, D. Collins, G. Li, *Geophys. Res. Lett.* **33**, L15802 (2006).
9. J. Zhao, A. Khalizov, R. Zhang, R. McGraw, *J. Phys. Chem. A* **113**, 680 (2009).
10. A. B. Nadykto, F. Yu, *Chem. Phys. Lett.* **435**, 14 (2007).
11. L. H. Young *et al.*, *Atmos. Chem. Phys.* **8**, 4997 (2008).
12. R. Zhang *et al.*, *Science* **304**, 1487 (2004).
13. T. Berndt *et al.*, *Science* **307**, 698 (2005).
14. J. Fan, R. Zhang, *Environ. Chem.* **1**, 140 (2004).
15. L. Wang *et al.*, *Nat. Geosci.* **3**, 238 (2010).
16. For nucleation involving sulfuric acid and species A, the nucleation rate, J , is expressed by $J = k[\text{H}_2\text{SO}_4]^{m_s} [\text{A}]^{m_A}$ where m_s is the number of sulfuric acid molecules in the critical cluster and k is a constant. The nucleation theorem is derived with the Gibbs's free energy reaching the maximum (ΔG^*) at the critical nucleus (r^*), relating the number of molecules, n_s , of the species, A, in the critical nucleus to the slope of the logarithm of the nucleation rate, as a function of the logarithm of the gaseous concentration of the nucleating species, [A], i.e., $n_A \approx \partial \ln J / \partial \ln [A]$.
17. Supported by the National Natural Science Foundation of China Grant (grant 40728006), the Robert A. Welch Foundation (grant A-1417), and the U.S. National Science Foundation (grants AGS-0938352 and CBET-0932705).

10.1126/science.1189732

The Transition Between Real and Complex Superconducting Order Parameter Phases in UPt_3

J. D. Strand,^{1*} D. J. Bahr,¹ D. J. Van Harlingen,^{1*} J. P. Davis,^{2†} W. J. Gannon,² W. P. Halperin²

Order parameter symmetry is one of the basic characteristics of a superconductor. The heavy fermion compound UPt_3 provides a rich system for studying the competition between superconductivity and other forms of electronic order and exhibits two distinct superconducting phases that are characterized by different symmetries. We fabricated a series of Josephson tunnel junctions on the as-grown surfaces of UPt_3 single crystals spanning the a - b plane. By measuring their critical current, we mapped out the magnitude of the superconducting order parameter as a function of the momentum-space direction and temperature. In the high-temperature phase, we observed a sharp node in the superconducting gap at 45° with respect to the a axis; an out-of-phase component appeared in the low-temperature phase, creating a complex order parameter.

The heavy fermion material UPt_3 is an unconventional superconductor with a distinct combination of features that have yet to be fully explained. It is one of only a very few known superconductors that have multiple superconducting phases. UPt_3 exhibits a double peak in the specific heat (I), indicating two distinct superconducting phases, with an initial transition at the upper critical temperature (T_{c+}) ~ 550 mK into the high-temperature A phase and a second transition at the lower critical temperature (T_{c-}) ~ 500 mK into the low-temperature B phase. In addition, a third phase has been identified at high magnetic fields (2). There is evidence for a very weak antiferromagnetic moment in the basal plane that coexists with superconductivity (3); pressure studies indicate that this finite moment may be responsible for the splitting of the superconducting transition (4). Transport measurements at low temperatures (5, 6) as well as anisotropic tunneling data (7) reveal the presence of nodes in the superconducting gap. There are indications from muon spin resonance of spontaneous magnetization and thus time-reversal symmetry breaking in the low-temperature phase (8, 9), and nuclear magnetic resonance studies of the Knight shift support a triplet-pairing mechanism or a singlet state with strong spin-orbit scattering (10).

A great deal of theoretical activity has been devoted to reconciling these disparate experimental phenomena within a single theory of the superconducting order (11, 12). Observations of the vortex lattice with small-angle neutron scattering (13, 14) and phase-sensitive measurements with Josephson interferometry (15) have been interpreted as being consistent with the E_{2u} representation of the order parameter (16). In this model, there is a real com-

ponent to the order parameter that turns on at T_{c+} and an imaginary component that turns on at T_{c-} , creating a complex order parameter state. The amplitude of the order parameter is described by

$$\Delta(T) = [(k_x^2 - k_y^2)k_z \Delta_R(T) + i(2k_x k_y)k_z \Delta_I(T)] \quad (1)$$

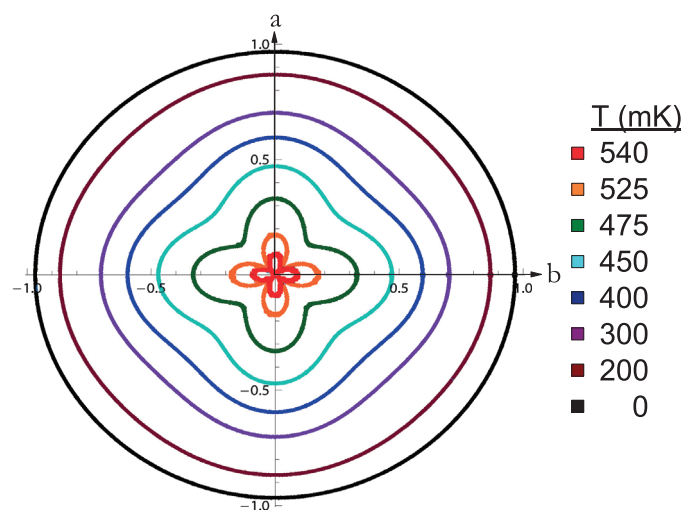
where $\Delta_R(T)$ and $\Delta_I(T)$ are the temperature-dependent magnitudes of the real and imaginary components. The evolution of the order parameter with temperature is shown in Fig. 1, in which we have assumed a simple temperature dependence of the form $\Delta_R(T) \sim 1 - (T/T_{c+})^2$ and $\Delta_I(T) \sim 1 - (T/T_{c-})^2$; we will show that these functional forms are consistent with the data. The primary feature is the series of nodes in the high-temperature phase located 45° off the gap maxima. These gradually fill in as the out-of-phase component grows in the low-temperature phase until the gap magnitude is isotropic in the zero-temperature limit. One particular aspect of the E_{2u} representation is that the fourfold symmetry of the order parameter differs from the hexagonal structure of the crystal

lattice. This suggests that the order parameter must select one of several degenerate configurations with respect to the crystal lattice. These configurations differ by a 30° rotation, and so the position of the nodes should in principle vary by $\pm 30^\circ$ with successive cooldowns of the crystal.

There are many measurements capable of identifying the initial transition from a normal metal to a superconductor, and it is also relatively straightforward to measure features of the combined energy gap by means of transport and spectroscopic probes that measure quasiparticle density. However, it is quite challenging to extract information about a second superconducting phase buried inside the first. The anisotropic shape of the energy gap in UPt_3 should allow one to separate the two phases, but to do so requires a high degree of momentum selectivity. Josephson junctions are an excellent candidate for this measurement because the tunneling probability drops off exponentially with barrier thickness so that a given junction effectively probes a single direction in k space. The critical current of a Josephson junction between two superconductors depends on the amplitude of the order parameters, and so a careful measurement of the critical current gives a direct measure of the order parameter for a given direction. This approach has been used successfully to map out the magnitude of the order parameter as a function of angle in the cuprates, demonstrating the d -wave symmetry (17). With this in mind, we fabricated Josephson junctions at a variety of angles around a single crystal of UPt_3 and measured the temperature dependence of the critical currents so as to map the evolution of the superconducting order parameter (18).

We report data from two samples (the first sample is shown in Fig. 2A). The junctions are spaced more or less evenly in angle over a 90° span between the a and b axes. For our second sample (Fig. 2B), we made a single wide junction centered at 45° from the a axis and then used a focused ion beam (FIB) to cut that junction into seven smaller junctions with different tunneling angles. This al-

Fig. 1. A two-dimensional representation of the superconducting order parameter, with the c axis out of the page. Each line represents the gap magnitude at a different temperature, with the innermost two lines above T_{c-} and the rest of the lines below T_{c-} . The nodes in the high-temperature phase gradually fill in as the low-temperature component of the gap grows. Not shown is a line node in the basal plane dividing the positive and negative c axis.



¹Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. ²Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208, USA.

*To whom correspondence should be addressed. E-mail: strand2@illinois.edu (J.D.S.); dvh@illinois.edu (D.J.V.H.)

†Present address: Department of Physics, University of Alberta, Edmonton, Alberta T6G 2G7, Canada.

lowed us to reduce the spacing between junctions and improve our angular resolution in the vicinity of the node. We measured the critical current as a function of temperature for each of our junctions by taking a series of current-voltage relation (I - V) plots (fig. S1), and we defined the temperature when the critical current of a junction dropped below our resolution of 0.1 μ A as the onset temperature of superconductivity, T_{onset} (Fig. 3A). We found that the onset of supercurrents lies just below the bulk superconducting transition T_{c+} for junctions at all angles except those very near 45° from the a axis ($\pm 3^\circ$). For those junctions, the onset of supercurrent

did not occur until near the lower transition, T_{c-} . The angular dependence of T_{onset} is shown in Fig. 3B. This confirms the location of a line node that is present in the gap of the high-temperature component of the order parameter but not the low-temperature component, providing further verification of the E_{2u} symmetry picture. For each sample, the data were reproducible over several thermal cycles, with a change in T_{onset} for a given junction of <1 mK.

We found that the location of the node never changes in successive cooldowns on the sample. It is somewhat surprising that the location of the node never changes because the order parameter should

be able to choose between several degenerate options during each cooldown. Both the observations reported here and previous experiments suggest that UPt_3 only forms a single superconducting domain (15, 19). It is possible that coupling to the anti-ferromagnetic moments in the lattice is providing a preferred orientation of the order parameter.

We also found that the temperature dependence of the critical current varied with the angle (Fig. 4A). The rate of increase of the critical current is slower the closer the junction is to the node at 45° , but there is always a finite supercurrent with an onset at T_{c+} unless the junction is at the node. For junctions close to the node, the rate of increase in critical current changes abruptly at T_{c-} , revealing the onset of the imaginary order parameter phase that increases the pairing amplitude and hence the critical current. We modeled this behavior with the E_{2u} representation of the gap magnitude from Eq. 1 assuming that $I_c(T) \sim \Delta(T)$. This model captures most of the features of our data using the temperature dependences assumed in Fig. 1, $\Delta_R(T) \sim 1 - (T/T_{c+})^2$ and $\Delta_i(T) \sim 1 - (T/T_{c-})^2$ (Fig. 4B). This shows that we can observe not just the presence or absence of a gap but also the angular dependence of the gap magnitude. In particular, there is a strong signature of the low-temperature component even at angles well away from the node (20).

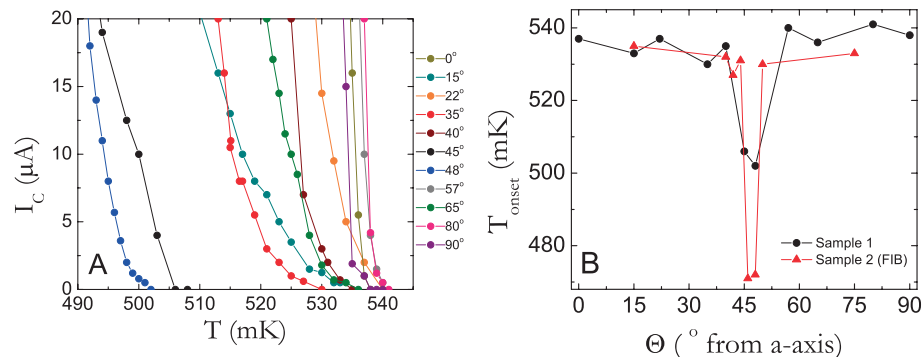
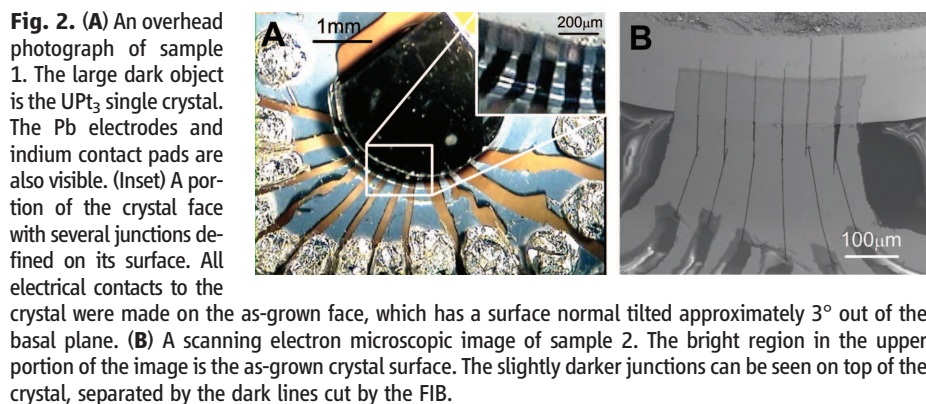


Fig. 3. (A) Plots of the critical current versus the temperature for Josephson junctions fabricated on the as-grown surface of sample 1. (B) The onset temperature for supercurrents versus the angle (measured from the a axis) of the junctions. The sharp dip at 45° indicates the presence of a node in the high-temperature component of the superconducting gap.

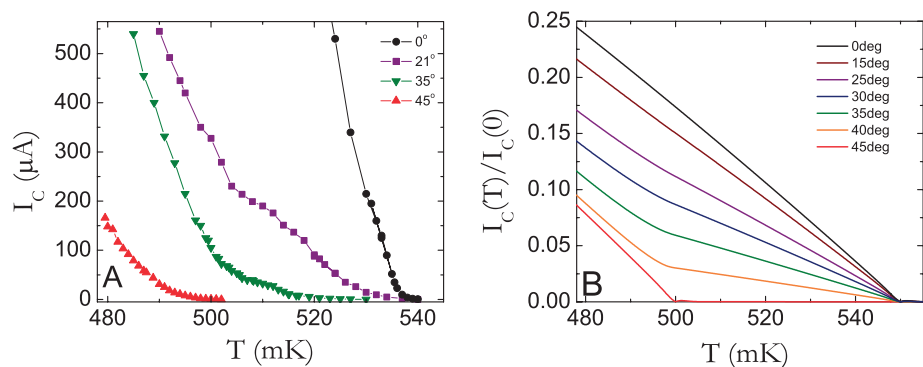


Fig. 4. (A) A series of $I_c(T)$ curves for junctions fabricated at different angles. The magnitude grows more slowly for greater angles, but the onset temperature remains the same for all angles other than 45° . A sharp upturn in the slope occurs at T_{c-} when the imaginary component of the order parameter begins. (B) A simulation of $I_c(T)$ curves for a combination of two gaps with different onset temperatures, one with nodes and one without.

References and Notes

1. R. A. Fisher *et al.*, *Phys. Rev. Lett.* **62**, 1411 (1989).
2. S. Adenwalla *et al.*, *Phys. Rev. Lett.* **65**, 2298 (1990).
3. G. Aeppli *et al.*, *Phys. Rev. Lett.* **60**, 615 (1988).
4. S. Hayden, L. Taillefer, C. Vettier, J. Flouquet, *Phys. Rev. B* **46**, 8675 (1992).
5. K. Behnia *et al.*, *J. Low Temp. Phys.* **84**, 261 (1991).
6. H. Suderow, J. P. Brison, A. Huxley, J. Flouquet, *J. Low Temp. Phys.* **108**, 11 (1997).
7. G. Goll, H. Löhneysen, I. K. Yanson, L. Taillefer, *Phys. Rev. Lett.* **70**, 2008 (1993).
8. G. M. Luke *et al.*, *Phys. Rev. Lett.* **71**, 1466 (1993).
9. P. D. de Reotier *et al.*, *Phys. Lett. A* **205**, 239 (1995).
10. H. Tou *et al.*, *Phys. Rev. Lett.* **77**, 1374 (1996).
11. W. C. Wu, R. Joynt, *Phys. Rev. B* **65**, 104502 (2002).
12. R. Joynt, L. Taillefer, *Rev. Mod. Phys.* **74**, 235 (2002).
13. A. Huxley, P. Rodiere, D. M. Paul, R. Cubitt, J. Flouquet, N. van Dijk, *Nature* **406**, 160 (2000).
14. T. Champel, V. P. Mineev, *Phys. Rev. Lett.* **86**, 4903 (2001).
15. J. D. Strand, D. J. Van Harlingen, J. B. Kycia, W. P. Halperin, *Phys. Rev. Lett.* **103**, 197002 (2009).
16. J. A. Sauls, *Adv. Phys.* **43**, 113 (1994).
17. H. J. H. Smilde *et al.*, *Phys. Rev. Lett.* **95**, 257001 (2005).
18. Materials and methods are available as supporting material on Science Online.
19. M. J. Graf, S. Yip, J. A. Sauls, *Phys. Rev. B* **62**, 14393 (2000).
20. Supporting text is available on Science Online.
21. The measurements were carried out at the University of Illinois at Urbana-Champaign in the Frederick Seitz Materials Research Laboratory and supported by NSF grant DMR07-05214. The UPt_3 crystals were grown and annealed at Northwestern University supported by the Department of Energy Basic Energy Sciences grant DE-FG02-05ER46248. We acknowledge the help of J. Pollanen, H. Choi, and T. Lippman with crystal growth and characterization.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1368/DC1

Materials and Methods

SOM Text

Figs. S1 and S2

References

4 February 2010; accepted 30 April 2010
10.1126/science.1187943

Mechanical Control of Spin States in Spin-1 Molecules and the Underscreened Kondo Effect

J. J. Parks,^{1,2} A. R. Champagne,^{1*} T. A. Costi,³ W. W. Shum,² A. N. Pasupathy,^{1†} E. Neuscamman,² S. Flores-Torres,² P. S. Cornaglia,⁴ A. A. Aligia,⁴ C. A. Balseiro,⁴ G. K.-L. Chan,² H. D. Abruña,² D. C. Ralph^{1,5‡}

The ability to make electrical contact to single molecules creates opportunities to examine fundamental processes governing electron flow on the smallest possible length scales. We report experiments in which we controllably stretched individual cobalt complexes having spin $S = 1$, while simultaneously measuring current flow through the molecule. The molecule's spin states and magnetic anisotropy were manipulated in the absence of a magnetic field by modification of the molecular symmetry. This control enabled quantitative studies of the underscreened Kondo effect, in which conduction electrons only partially compensate the molecular spin. Our findings demonstrate a mechanism of spin control in single-molecule devices and establish that they can serve as model systems for making precision tests of correlated-electron theories.

The electronic states of atoms and molecules depend on the symmetry imposed by their local environment. A simple manifestation occurs in the attachment of ligands to a transition-metal ion to form a coordination complex, which breaks spherical symmetry and causes splittings within the ion's initially degenerate d orbitals. For a complex having spin $S \geq 1$, additional distortions of the ligands combined with spin-orbit coupling can cause splittings within the initially $(2S + 1)$ -degenerate spin states, giving rise to magnetic anisotropy (1, 2). To study the effects of symmetry-breaking distortions, we stretched individual $S = 1$ molecules within mechanically controllable devices (3). Simultaneous electron transport measurements showed that stretching lifts the degeneracy of the $S = 1$ ground state and enables control of the magnetic anisotropy. The same devices also enabled tests of predictions (4) for the temperature dependence of the underscreened $S = 1$ Kondo effect (5–7).

We studied the $\text{Co}(\text{tpy-SH})_2$ complex (where tpy-SH is 4'-mercapto-2,2':6',2''-terpyridine) (Fig. 1A), in which a Co ion resides in an environment of approximate octahedral symmetry, through its coordination to six N atoms on two terpyridine ligands. When attached to gold electrodes and cooled to low temperature, this molecule exhibits

the Kondo effect (8); the spin of the molecule is screened by electron spins in the electrodes, leading to a peak in the electrical conductance at zero bias voltage, V (9). We used the Kondo effect as a spectroscopic probe to interrogate the molecular spin (10). Previous electron-transport studies of individual metal complexes have probed rich behavior, including not only Kondo physics (8, 11–13) but also vibrational excitations (13–15) and molecular magnetism (16, 17).

We used mechanically controllable break-junction devices (Fig. 1B) to stretch individual molecules while measuring their conductance. The devices were made by fabricating a Au wire suspended above a thin Si substrate (Fig. 1C), depositing molecules, and then using electromigration (18) to create a molecular-scale break in the wires before beginning studies in which we

tuned the distance between electrodes by bending the substrate. Details, statistics, and discussion of control experiments are provided in (19). We have also measured the same $\text{Co}(\text{tpy-SH})_2$ complex by using fixed Au electrodes (8, 11, 13, 15–17).

Measurements of differential conductance (dI/dV) as a function of increasing electrode spacing at a temperature $T = 1.6$ K are shown in Fig. 2. At the initial position of the electrodes, the dI/dV spectra exhibited a single peak centered at $V = 0$ with amplitude of order (but less than) $2e^2/h$, a signature of Kondo-assisted tunneling through the molecule. As we stretched each molecule, the single conductance peak split into two beyond a value for the change in electrode spacing that varied from device to device. For the two devices displayed in Fig. 2 and for two others [supporting online material (SOM) section S2], we could reproducibly cross this transition back and forth between one peak and two. For the stretched molecules (Fig. 2E), the temperature dependence of dI/dV at $V = 0$ showed a nonmonotonic dependence similar to the V dependence upon increasing $|V|$ from 0.

The observed splitting of the Kondo peak as a function of stretching is in contrast to a previous study of the spin- $1/2$ Kondo effect in C_{60} molecules (20), in which varying the electrode spacing modified the height and width of the Kondo resonance but did not split the peak. We show that the peak splitting for the $\text{Co}(\text{tpy-SH})_2$ complex is caused by a higher-spin $S = 1$ Kondo effect, together with the breaking of degeneracy within the $S = 1$ triplet ground state caused by molecular distortion (2). For an unstretched $S = 1$ ion in a ligand field with octahedral symmetry, the triplet states are strictly degenerate according to group theory. However, if the molecule is stretched axially (the z axis), the $S_z = 0$ state will be lowered by a zero-field splitting energy D be-

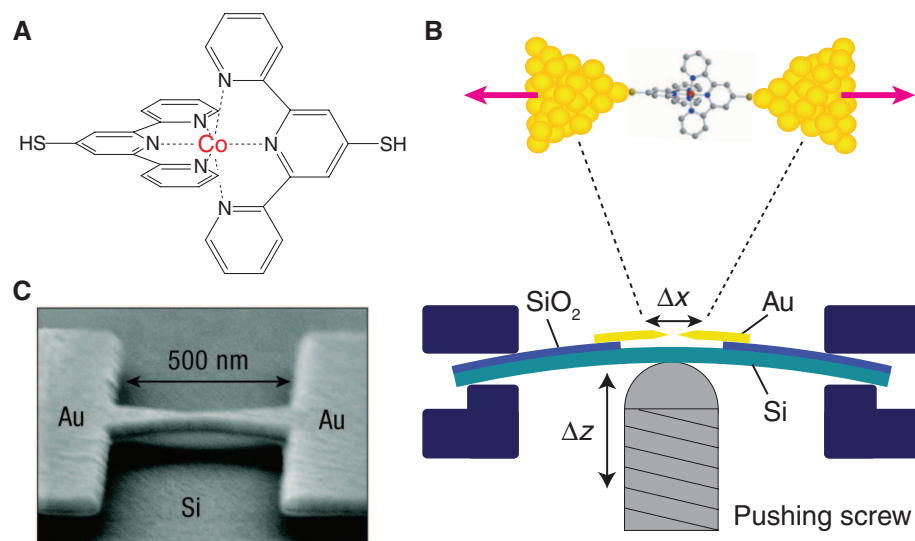


Fig. 1. Single-molecule electrical devices based on a transition-metal complex. (A) Chemical structure of $\text{Co}(\text{tpy-SH})_2$, where tpy-SH is 4'-mercapto-2,2':6',2''-terpyridine. (B) Schematic of the mechanically controllable break junction. (C) Scanning electron micrograph of a suspended break junction before molecule deposition.

¹Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA. ²Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA. ³Institut für Festkörperforschung and Institute for Advanced Simulation, Forschungszentrum Jülich, 52425 Jülich, Germany. ⁴Centro Atómico Bariloche and Instituto Balseiro, Comisión Nacional de Energía Atómica (CNEA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), 8400 San Carlos de Bariloche, Argentina. ⁵Kavli Institute at Cornell, Cornell University, Ithaca, NY 14853, USA.

*Present address: Department of Physics, Concordia University, Montréal, Québec H4B 1R6, Canada.

†Present address: Department of Physics, Columbia University, New York, NY 10027, USA.

‡To whom correspondence should be addressed. E-mail: ralph@ccmr.cornell.edu

low the $S_z = \pm 1$ states, corresponding to a uniaxial spin anisotropy (Fig. 2F). This broken degeneracy quenches the Kondo resonance near $V = 0$ and causes conductance peaks at $V = \pm D/e$ because of inelastic tunneling. The distortion-induced breaking of the ground-state degeneracy for an $S = 1$ molecule is in contrast to the physics of half-integer spins ($2I$) for which time-reversal symmetry mandates that the ground state remains a degenerate Kramers doublet. To establish this picture of stretching-induced control of spin states, we show that the spin of our molecules is indeed $S = 1$, on the basis of temperature and magnetic-field studies, and we verify that stretching produces spin anisotropy by measuring how the level splitting depends on the direction of an applied magnetic field.

The temperature dependence of the Kondo conductance is predicted to depend on the molecular spin, S , and the number of screening channels in the electrodes. To be fully screened at zero temperature, a molecule with spin S requires coupling to $2S$ screening channels (5). In our experimental geometry, there are two screening channels consisting of linear combinations of states from the two electrodes with different couplings J_1 and J_2 to the molecular spin that result in two Kondo temperature scales, T_{K1} and T_{K2} (22) (we assume $T_{K1} < T_{K2}$). The Kondo temperatures depend exponentially on the couplings, so in the typical case T_{K1} is much smaller than T_{K2} . If $S > 1/2$, the result is that over the range $T_{K1} \ll T < T_{K2}$ the original spin is only partially screened, to a value $S - 1/2$. Relative to the fully screened $S = 1/2$ Kondo effect, this underscreened effect should produce a much slower rise in the conductance as T decreases below T_{K2} ($4, 6, 7$). Henceforth we denote T_{K2} as simply T_K .

Figure 3, A and B, shows the measured zero-bias conductance $G(T)$ for the unstretched configuration of devices A and B with fits to numerical renormalization group (NRG) predictions for the fully screened $S = 1/2$ Kondo model ($23, 24$) and the underscreened $S = 1$ and $3/2$ models (4) (for fitting details, see SOM S3). Each fit has two adjustable param-

eters, T_K and the zero-temperature conductance $G(0)$. We find that $G(T)$ deviates strongly from the form for the $S = 1/2$ Kondo effect

and instead agrees quantitatively with the prediction for the $S = 1$ underscreened case. We performed the same analysis for 10 unstretched

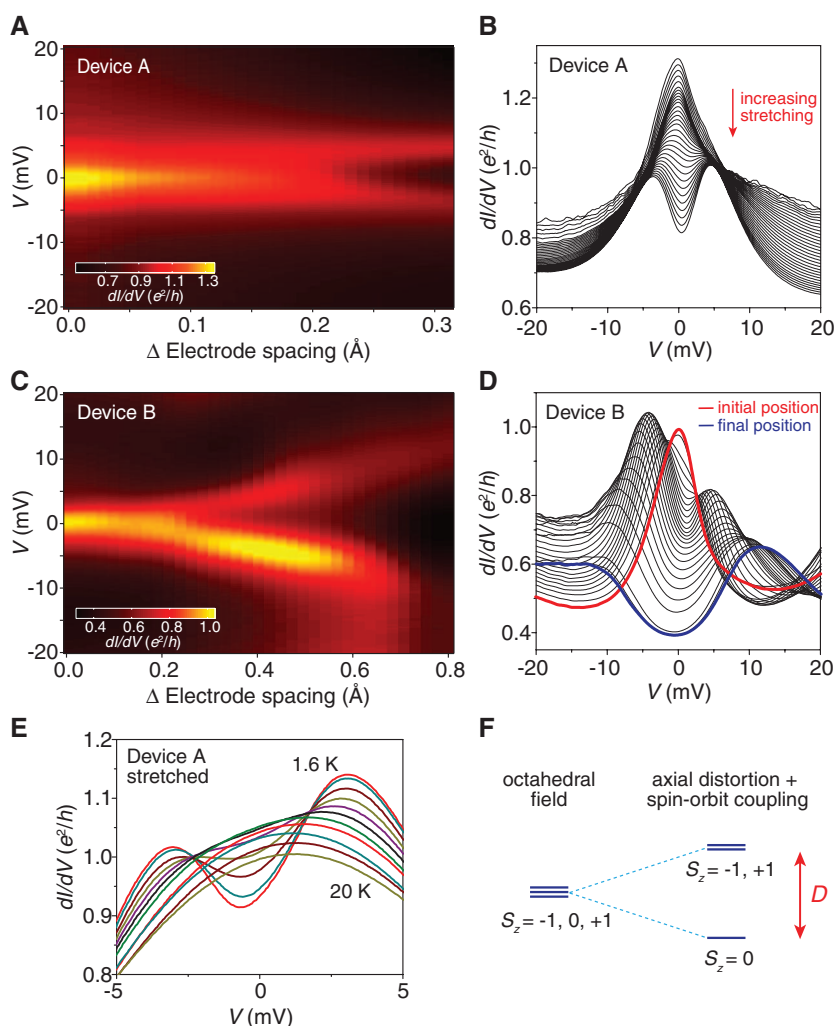


Fig. 2. Splitting of the Kondo peaks as a function of mechanical stretching. (A and C) dI/dV as a function of bias voltage and electrode spacing at $T = 1.6$ K for (A) device A and (C) device B. (B and D) Line cuts showing dI/dV as a function of V for different values of molecular stretching for (B) device A and (D) device B. (E) Temperature dependence of the conductance for device A on the stretched side of the splitting transition. (F) Breaking of the degeneracy among the $S = 1$ triplet states by uniaxial distortion.

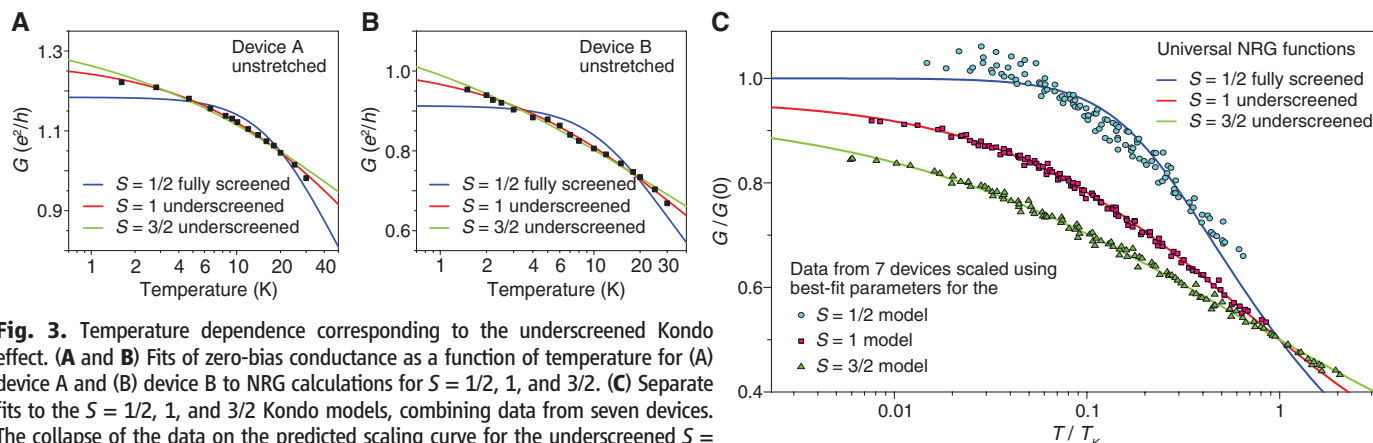


Fig. 3. Temperature dependence corresponding to the underscreened Kondo effect. (A and B) Fits of zero-bias conductance as a function of temperature for (A) device A and (B) device B to NRG calculations for $S = 1/2, 1$, and $3/2$. (C) Separate fits to the $S = 1/2, 1$, and $3/2$ Kondo models, combining data from seven devices. The collapse of the data on the predicted scaling curve for the underscreened $S = 1$ Kondo model is excellent; the normalized root mean square scatter for the $S = 3/2$ ansatz is greater by 44% and for $S = 1/2$ by 320%. For the $S = 1$ fits, the seven Kondo temperatures are 1.1, 28, 62, 81, 84, 180, and 210 K.

Co(tpy-SH)₂ devices (SOM S4). Seven showed similarly unambiguous agreement with underscreened Kondo scaling, and all of these gave superior fits to the $S = 1$ prediction than to $S = 3/2$. Discussion of the other three samples is provided in SOM S4.

In Fig. 3C, for the seven devices with underscreened characteristics, we plotted $G(T)/G(0)$ versus T/T_K by using the parameters $G(0)$ and T_K obtained from separate fits to the $S = 1/2$, 1, and $3/2$ models. This allowed us to test how well the data can all be described by the predicted scaling

curves. We used the root mean square deviation of the data from each theory curve normalized to the average of the scaled data as a goodness-of-fit metric. The $S = 1$ fit is best: The deviation of points for the $S = 1/2$ ansatz is greater by 320%, for $S = 3/2$ by 44%, and for larger S by more than 80%. Additional confirmation that $S = 1$ is discussed below based on magnetic field studies. An independent (25) observation of the temperature scaling predicted for the underscreened Kondo effect has also been reported recently for a single $S = 1$ C₆₀ device (26).

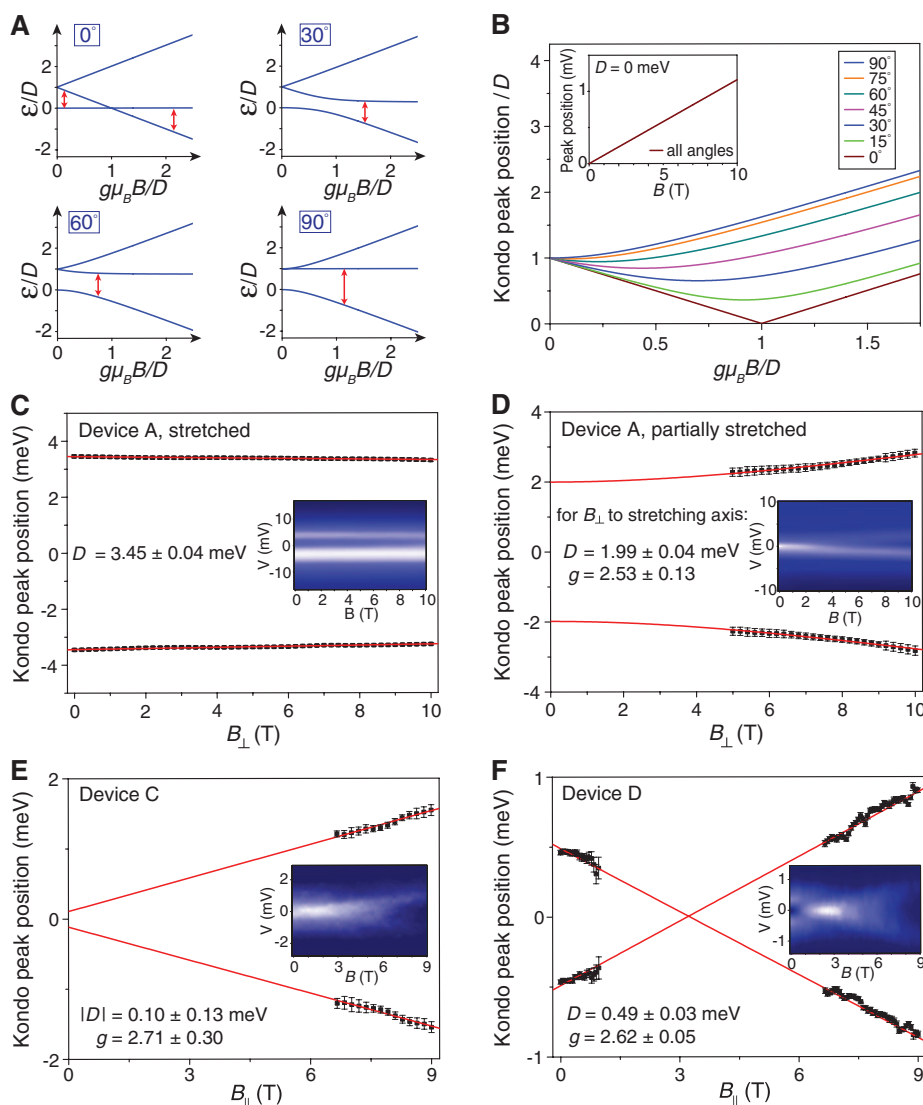


Fig. 4. The magnetic field evolution of the Kondo peaks demonstrates the presence of spin anisotropy. (A) Energy eigenvalues (for $S = 1$) of the model spin-anisotropy Hamiltonian $H = -g\mu_B \mathbf{B} \cdot \mathbf{S} + DS_z^2$ for four different field angles with respect to the anisotropy axis. The red arrows indicate the lowest-energy inelastic transitions corresponding to the finite-bias Kondo peaks. (B) Predicted Kondo peak position versus magnetic field at several field angles, with energies normalized to the zero-field splitting. (Inset) Kondo peak splitting (for $g = 2$) as a function of magnetic field for all angles in the absence of spin anisotropy. (C and D) Kondo peak positions versus magnetic field at $T = 1.6$ K in device A for two different values of stretching with the magnetic field oriented approximately perpendicular to the molecule axis. The data in (D) can be fit for any field angle between 70° and 90° , giving parameters in the range $D = 2.0$ to 2.2 meV and $g = 2.5$ to 2.8 . (E and F) The magnetic field dependence in fixed-electrode devices C and D with the magnetic field approximately along the molecular axis. (Insets) Color plots of dI/dV versus bias voltage and magnetic field.

Our interpretation that the stretching-induced Kondo splitting is caused by the breaking of spin degeneracy is confirmed by the presence of spin anisotropy in the stretched state: the magnetic field (B) dependence of the Kondo splitting differs depending on the angle at which B is applied relative to the stretching axis. Figure 4A shows the expected B dependences of the energy levels in an $S = 1$ multiplet for $D > 0$ and for several field angles, and Fig. 4B displays the energy differences between the first excited state and the ground state, which determine the Kondo peak positions. For a field orientation approximately perpendicular to the anisotropy axis ($\theta = 90^\circ$), when $g\mu_B B \ll D$ (where μ_B is the Bohr magneton) the B dependence of the peak position should be very weak, whereas when $g\mu_B B$ grows large enough to be comparable to D , the peak should shift gradually to larger values of $|V|$ with positive curvature. For B parallel to the anisotropy axis ($\theta = 0^\circ$), the B dependence of the peak position should be linear and much stronger. Our measurements show a dependence on the orientation of B just as expected within this model. Figure 4, C and D, shows the B dependence for different degrees of stretching for device A, with $\theta \approx 90^\circ$. At a relatively large degree of stretching (Fig. 4C), $D = 3.45$ meV was larger than $g\mu_B B$ at our largest field (~ 1.5 meV), and the Kondo peak positions were almost independent of B . The slightly negative slope of $d|V|/dB$ [corresponding to an effective g factor of 0.21 ± 0.01 (SD)] suggests that the stretching axis is not exactly perpendicular to B . For a smaller degree of stretching (Fig. 4D), $g\mu_B B$ was comparable to D at large fields, and the Kondo peaks shifted weakly to larger $|V|$ as a function of B with positive curvature. Figure 4, E and F, shows data for B approximately parallel to the molecular axis for fixed-electrode devices, which exhibit zero-field splitting without intentional mechanical stretching (among devices that had Kondo features, $\sim 10\%$ of the fixed-electrode devices and $\sim 20\%$ of the adjustable devices exhibited a split Kondo peak at the initial electrode spacing, suggesting that these molecules happened to be strained initially). In Fig. 4, E and F, the Kondo peaks shift linearly with B , with much larger slopes of $d|V|/dB$ than for $\theta \approx 90^\circ$, corresponding to g factors of 2.6 to 2.7 (± 0.3). In device D, the peaks shift initially to smaller $|V|$, pass through zero, and then move to larger $|V|$ (Fig. 4F). From the negative slopes of $d|V|/dB$ in Fig. 4, C and F, we can confirm that $D > 0$. The scale of splittings we measure, ~ 0 to 5 meV, agrees with our simulations (SOM S6) and is typical of zero-field splittings in distorted coordination complexes (2).

These magnetic field data provide further evidence that the spin state of the molecule is $S = 1$, rather than $1/2$, $3/2$, or any half integer. For half-integer spins with $D > 0$, the ground state for $B = 0$ is the degenerate Kramers doublet $S_z = \pm 1/2$. Regardless of the direction of B or the degree of stretching, this doublet will give rise to a Kondo peak that shifts approximately linearly with B and

extrapolates to zero splitting at $B = 0$ (2I), in contradiction to the data (SOM S7). Integer values of $S \geq 2$ can be ruled out on the basis of the temperature scaling described above and because previous studies of sixfold coordinated Co complexes have found almost exclusively $S \leq 3/2$, with exceptions only for fluoride ligands (I, 2). We conclude that only $S = 1$ can explain the measurements, and from this we identify the charge state of the metal center to be Co^{1+} (SOM S8).

In coordination chemistry, the existence of zero-field splittings induced by molecular distortion is well established, but the ability we demonstrate to continuously distort an individual molecule while simultaneously measuring its zero-field splitting opens the possibility for dramatically more detailed and precise comparisons with theory. For correlated-electron physics, our results demonstrate that single-molecule electrical devices can provide well-controlled model systems for studying $S \geq 1$ underscreened Kondo effects not previously realizable in experiment. Our work further demonstrates that mechanical control can be a realistic strategy for manipulating molecular spin states to supplement or replace the use of magnetic fields in proposed applications such as quantum manipulation or information storage (27, 28).

References and Notes

1. A. Abragam, B. Bleaney, *Electron Paramagnetic Resonance of Transition Ions* (Dover, New York, 1986).
2. R. Boca, *Coord. Chem. Rev.* **248**, 757 (2004).
3. N. Agraït, A. L. Yeyati, J. M. van Ruitenbeek, *Phys. Rep.* **377**, 81 (2003).
4. F. Mallet et al., *Phys. Rev. Lett.* **97**, 226804 (2006).
5. P. Nozières, A. Blandin, *J. Phys. (Paris)* **41**, 193 (1980).
6. A. Posazhennikova, P. Coleman, *Phys. Rev. Lett.* **94**, 036802 (2005).
7. P. Mehta, N. Andrei, P. Coleman, L. Borda, G. Zarand, *Phys. Rev. B* **72**, 014430 (2005).
8. J. Park et al., *Nature* **417**, 722 (2002).
9. M. Grobis, I. G. Rau, R. M. Potok, D. Goldhaber-Gordon, in *Handbook of Magnetism and Advanced Magnetic Materials*, H. Kronmüller, S. S. P. Parkin, Eds. (Wiley, Hoboken, NJ, 2007).
10. C. Romeike, M. R. Wegewijs, W. Hofstetter, H. Schoeller, *Phys. Rev. Lett.* **97**, 206601 (2006).
11. W. Liang, M. P. Shores, M. Bockrath, J. R. Long, H. Park, *Nature* **417**, 725 (2002).
12. A. Zhao et al., *Science* **309**, 1542 (2005).
13. L. H. Yu et al., *Phys. Rev. Lett.* **93**, 266802 (2004).
14. X. H. Qiu, G. V. Nazin, W. Ho, *Phys. Rev. Lett.* **92**, 206102 (2004).
15. D.-H. Chae et al., *Nano Lett.* **6**, 165 (2006).
16. H. B. Heersche et al., *Phys. Rev. Lett.* **96**, 206801 (2006).
17. M.-H. Jo et al., *Nano Lett.* **6**, 2014 (2006).
18. H. Park, A. K. L. Lim, A. P. Alivisatos, J. Park, P. L. McEuen, *Appl. Phys. Lett.* **75**, 301 (1999).
19. Materials and methods are available as supporting material on Science Online.
20. J. J. Parks et al., *Phys. Rev. Lett.* **99**, 026601 (2007).
21. A. F. Otte et al., *Nat. Phys.* **4**, 847 (2008).
22. M. Pustilnik, L. I. Glazman, *Phys. Rev. Lett.* **87**, 216601 (2001).
23. T. A. Costi, A. C. Hewson, V. Zlatić, *J. Phys. Cond. Matter* **6**, 2519 (1994).
24. D. Goldhaber-Gordon et al., *Phys. Rev. Lett.* **81**, 5225 (1998).
25. J. J. Parks, thesis, Cornell University, Ithaca, NY (2009).
26. N. Roch, S. Florens, T. A. Costi, W. Wernsdorfer, F. Balestro, *Phys. Rev. Lett.* **103**, 197202 (2009).
27. G. Christou, D. Gatteschi, D. N. Hendrickson, R. Sessoli, *MRS Bull.* **25**, 66 (November 2000).
28. L. Bogani, W. Wernsdorfer, *Nat. Mater.* **7**, 179 (2008).
29. We thank I. Cohen, M. Grobis, G. Hutchison, and P. McEuen for discussions and K. Bolotin, J. Grose, F. Kuemmeth, and E. Tam for technical help. Research at Cornell was supported by the NSF through the Cornell Center for Materials Research, DMR-0605742, CHE-0403806, and use of the Cornell Nanofabrication Facility/National Nanotechnology Infrastructure Network. T.A.C. acknowledges supercomputer support by the John von Neumann Institute for Computing (Jülich). P.S.C., A.A.A., and C.A.B. were supported by Proyectos de Investigación Plurianuales 11220080101821 of CONICET.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1370/DC1
Materials and Methods

SOM Text
Figs. S1 to S8
Table S1
References

11 January 2010; accepted 29 April 2010
10.1126/science.1186874

Nanoscale Tunable Reduction of Graphene Oxide for Graphene Electronics

Zhongqing Wei,^{1*} Debin Wang,^{2*} Suenne Kim,² Soo-Young Kim,^{3,4} Yike Hu,² Michael K. Yakes,¹ Arnaldo R. Laracuente,¹ Zhenting Dai,⁵ Seth R. Marder,³ Claire Berger,^{2,6} William P. King,⁵ Walter A. de Heer,² Paul E. Sheehan,^{1†} Elisa Riedo^{2†}

The reduced form of graphene oxide (GO) is an attractive alternative to graphene for producing large-scale flexible conductors and for creating devices that require an electronic gap. We report on a means to tune the topographical and electrical properties of reduced GO (rGO) with nanoscopic resolution by local thermal reduction of GO with a heated atomic force microscope tip. The rGO regions are up to four orders of magnitude more conductive than pristine GO. No sign of tip wear or sample tearing was observed. Variably conductive nanoribbons with dimensions down to 12 nanometers could be produced in oxidized epitaxial graphene films in a single step that is clean, rapid, and reliable.

Graphene's high electronic mobility (I) has been harnessed in devices such as transistors operating at gigahertz frequency (2); however, the zero band gap of graphene leads to high leakage currents in many applications. Another interesting material for a range of applications is graphene oxide (GO) (3, 4), which exhibits a transport gap greater than 0.5 eV at room temperature and becomes a semiconductor or semimetal as it is reduced back toward graphene (5, 6). Reduced GO (rGO) resembles graphene but with some residual oxygen and structural defects, yielding a conductivity that is comparable to that of doped conductive polymers (7) and 33,000 times higher than that of doped hydrogenated

amorphous Si (8). Reduced GO can also be used in highly sensitive gas sensors (9) and mechanical resonators with figures of merit surpassing those of graphene resonators (10).

We present a tip-based thermochemical nanolithography method to control the extent of reduction of GO and pattern nanoscale regions of rGO within a GO sheet at speeds of several $\mu\text{m/s}$. The relative increase in conductivity is as high as four orders of magnitude. GO was converted to rGO with a 100% yield in dozens of structures patterned on random locations in the GO film. Reduced GO patterns range from ribbons 12 nm in width (full width at half maximum, FWHM) up to 20 μm . No sign of tip wear or sample tearing

was observed, indicating that the "carbon skeleton" is continuous across the GO/rGO junction. Thermochemical nanolithography (TCNL) with heated probe tips can localize thermally induced chemical reactions on a surface (11–14) or deposit material (15–17). Similar heated tips have also been used to mechanically modify a polymer film (18). We performed TCNL by using a heated atomic force microscope (AFM) probe tip to reduce selected regions of both single layers of isolated GO and large-area GO films formed by on-chip oxidation of epitaxial graphene (GO_{epi}) grown on SiC.

Exposure of GO to strong reducing agents like hydrazine results in an increased electrical conductivity by three to four orders of magnitude (19). Thermal reduction of GO occurs already at moderate temperature (100° to 250°C) and enables tuning the gap in graphene oxide (6), as demonstrated in its current-voltage (I - V) characteristics. Recent studies have shown that annealing

¹Chemistry Division, U.S. Naval Research Laboratory, Code 6177, Washington, DC 20375, USA. ²School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA.

³School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA. ⁴School of Chemical Engineering and Materials Science, Chung Ang University, Seoul 156-756, Republic of Korea. ⁵Department of Mechanical Science and Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA. ⁶CNRS-Institut Néel, BP166, 38042 Grenoble Cedex 9, France.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: paul.sheehan@nrl.navy.mil (P.E.S.); elisa.riedo@physics.gatech.edu (E.R.)

GO at 450°C or above is equivalent to chemical reduction via hydrazine monohydrate at 80°C followed by heating at 200°C (20). We verified TCNL reduction of GO by friction force microscopy (FFM), conductive AFM (CAFM), Raman spectroscopy, Kelvin probe force microscopy (KPFM), and ultrahigh-vacuum (UHV) electronic transport measurements using a two- and four-point probe scanning tunneling microscope (STM) [see details in the supporting online material (SOM) (21)].

Arbitrary rGO features such as a cross (Fig. 1) or squares (Fig. 2) are reliably obtained by scanning the heated AFM tip over isolated GO flakes on a SiO₂/Si substrate. The thermal reduction decreases the 9.5 ± 1.9 Å height of the sheet by 2 to 5 Å, as obtained from the topography image (Fig. 1 and fig. S6). Two effects could lead to height reduction. One is the loss of oxygen-rich functional groups from the GO flake surface. Given that scanning an unheated tip does not result in height changes, this loss is primarily caused by intrinsic chemical conversion rather than mechanical removal. It is not possible, however, to rule out tribochemical effects at elevated temperatures. Second, the conversion of GO's sp³ carbon bonds into sp² carbon bonds will flatten the material because the sp³ carbon bonds in GO ripple the carbon skeleton, thereby increasing the sheet thickness (22–24).

Friction measurements show that variable reduction of GO could be achieved by controlling the temperature of the AFM tip. Graphene has a low friction coefficient (25), whereas oxides typically have higher friction coefficients. Thermal reduction should also reduce friction as the high-friction GO is replaced with lower-friction graphene. Figure 2 shows the strong correlation between the cantilever temperature during TCNL processing and the lateral force on a room-temperature tip scanned over previously reduced squares. Whereas the cantilever temperature can be precisely determined, the contact temperature must be modeled (see discussion in the SOM) (21); the reported temperatures are the cantilever temperatures. Reduction begins at or above 130°C, which is comparable to the results of Wu *et al.* and Mattevi *et al.*, who showed that reduction starts at 100°C (6, 20), presumably after the desorption of adventitious water. Higher temperatures increased the rate of reduction, as shown by the roughly linear decrease in relative friction with temperature.

Although isolated GO flakes are suited for basic studies, further technological development requires extended films of GO. Large-area GO_{epi} films (>15 mm²) were obtained by oxidizing multilayer epitaxial graphene (EG) grown on the carbon face of SiC [see material details in the SOM (21)]. The oxidized films consist of multiple high-quality GO_{epi} layers that completely cover the SiC surface. AFM images show no tearing in the GO_{epi} films, indicating that they maintain their structural integrity when exposed to the harsh oxidation conditions. Figures 3 and 4 show the re-

Fig. 1. Local thermal reduction of a single-layered graphene oxide flake. (A) Topography of a cross shape of reduced GO formed after an AFM tip heats the contact to 330°C scanned across the GO sheet at 2 μm/s. (B) The averaged profile of the trench outlined in (A) shows that the width (FWHM) of the line can be as narrow as 25 nm.

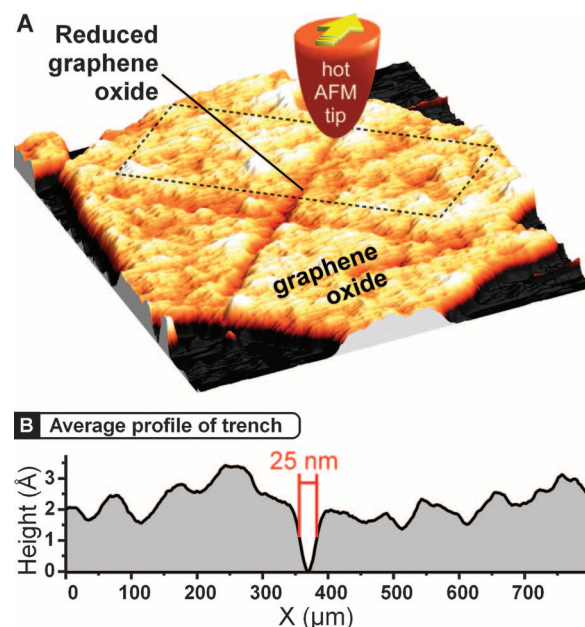
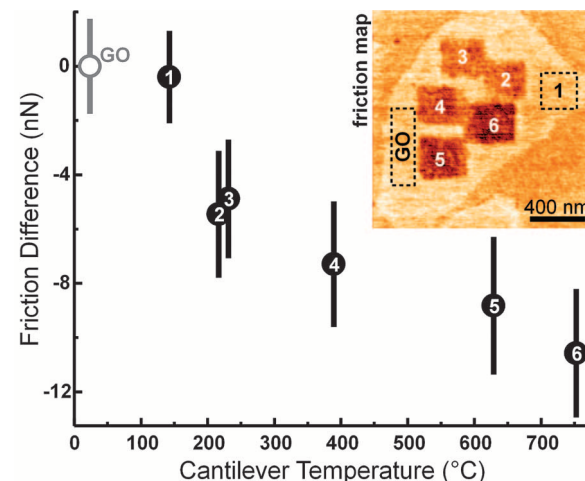


Fig. 2. The rate of thermal reduction depends on the tip temperature. The plot shows the decrease in lateral force on an AFM tip at room temperature as it scans over several squares previously reduced by TCNL at different temperatures. The inset is a room-temperature friction image of the GO sheet on which a heated tip was previously rastered twice over six square areas, at a speed of 4 μm/s. In square 1, the tip was heated during TCNL to $T_{\text{heater}} \sim 100^\circ\text{C}$, yielding no apparent reduction, whereas at temperatures $T_{\text{heater}} > 150^\circ\text{C}$ the rastered areas (squares 2 to 6) were thermally reduced. Reduced GO, which like bulk graphite behaves as a lubricant, shows lower friction than the original GO. Higher temperatures accelerate the thermal reduction of GO and thereby more rapidly lower friction.



sults obtained by performing TCNL on GO_{epi} films with different thicknesses, as determined by AFM by scratching away GO_{epi} from the SiC substrate [see the SOM (21)]. Figure 3 presents a zigzag rGO_{epi} nanoribbon written with a single line scan at $T_{\text{heater}} \sim 1060^\circ\text{C}$ on GO_{epi}. Figure 3A is an image of the current measured between a conductive platinum AFM tip and each point of the surface, showing no current on the GO surface and a current enhancement of about 100 pA in the rGO_{epi} nanoribbons. These current values are consistent with the presence of 12-nm-wide and several-nanometers-thick rGO_{epi} nanoribbons, presenting a vanishingly small Schottky barrier, and a resistive SiC substrate (resistivity of about 10^5 ohm-cm). For a 25-nm-thick GO film locally heated by a tip at 1000°C, heat flow through the layers might reduce most of the GO underneath the tip and leave only a few layers of

GO at the SiC interface, as shown in the SOM (21). The topographical image (Fig. 3B and black graph in Fig. 3C) indicates that the reduction produces a shallow indentation of 1 nm whose origin has been previously discussed for the isolated GO sheets.

We further investigated the electrical properties of the locally reduced GO_{epi} structures using KPFM and four-point probe transport measurements in a UHV Omicron Nanoprobe system [see details in the SOM (21)]. The sheet resistance, R_{sheet} , of 20 μm by 20 μm squares of TCNL rGO_{epi} decreased with increasing temperature used for the TCNL local reduction, up to four orders of magnitude lower than the resistance of the original GO_{epi} (427 ± 11 megohm) (6). The same decrease of the in-plane resistivity was observed for extended films of rGO_{epi} produced by overnight heating of GO_{epi} in a furnace at 600°C ($18 \pm$

Fig. 3. Local thermal reduction of a GO_{epi} film: current and topographical images. **(A)** Room-temperature AFM current image (taken with a bias voltage of 2.5 V between tip and substrate) of a zigzag-shaped nanoribbon fabricated by TCNL on GO_{epi} at $T_{\text{heater}} \sim 1060^\circ\text{C}$ with a linear speed of $0.2 \mu\text{m s}^{-1}$ and a load of 120 nN. **(B)** Corresponding topography image taken simultaneously with (A). **(C)** Averaged profiles of current and height of the cross sections that are indicated as dashed lines in (A) and (B).

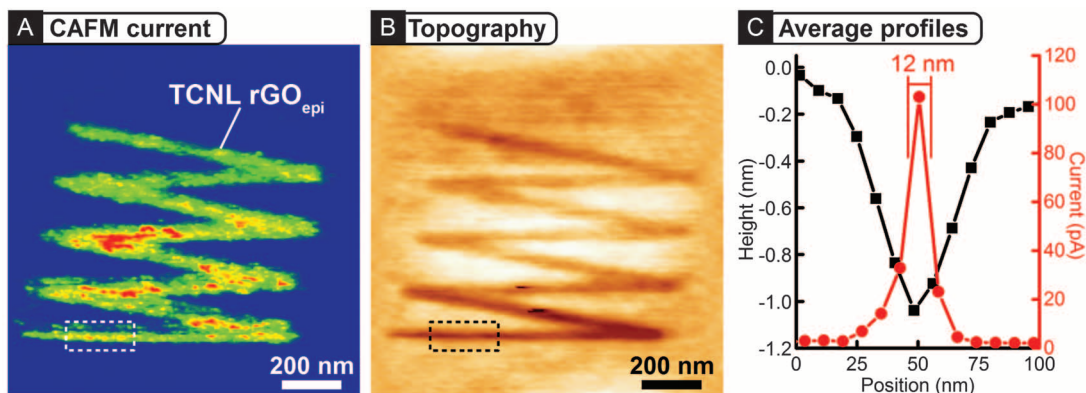
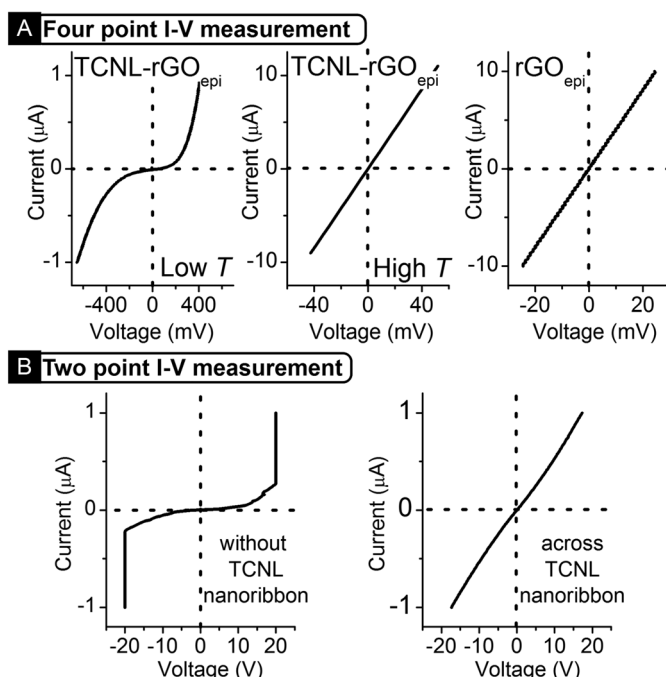


Fig. 4. Four-point and two-point transport measurements. **(A)** I - V curves obtained by four-point transport measurements of TCNL-reduced graphene oxide squares reduced at low temperature (Low T , $T_{\text{heater}} \sim 600^\circ\text{C}$), TCNL-reduced graphene oxide squares reduced at high temperature (High T , $T_{\text{heater}} \sim 1200^\circ\text{C}$), and furnace-reduced graphene oxide at 600°C in vacuum. **(B)** I - V curves obtained by two-point transport measurements of current between two rGO_{epi} squares with no nanoribbons in between (left curve), and between two rGO_{epi} squares with a nanoribbon in between (right curve).



10 kilohm). Furthermore, R_{sheet} and the shape of the I - V characteristics could be varied by changing the temperature of the AFM probe (in Fig. 4A, $R_{\text{sheet}} = 9174$ kilohm and 30 kilohm for low and high temperature, respectively). Kelvin probe measurements show that TCNL rGO_{epi} displays a contact potential change of 168 ± 54 mV in respect to GO_{epi} , similar to that of bulk reduced rGO (188 ± 96 mV). The presence of residual oxygen and structural disorder led to the large difference in conductivity between epitaxial graphene and rGO_{epi} or TCNL- rGO_{epi} .

We also analyzed an isolated TCNL- rGO_{epi} nanoribbon (Fig. 4B) with a length of $25 \mu\text{m}$ and a width of 100 nm, as measured by AFM. We acquired I - V data by placing conductive tips on top of two micrometer-sized squares of rGO_{epi} fabricated in situ by an electron beam at each end of the nanoribbon. Two-point transport measurements indicated a resistance larger than 2 gigohm when the tips were placed at an arbitrary position on the GO surface (very large

barrier at the contact) and a drop in resistance from 120 megohm (between the two squares with no nanoribbon) to 20 megohm (between the two squares connected by the nanoribbon). The transport changed from insulating to metallic (i.e., linear I - V curves) in the presence of the TCNL- rGO_{epi} nanoribbon between the squares (Fig. 4B). By using the relation $R_{\text{sheet}} = (R_{\text{ribbon}} \cdot w \cdot t_{\text{ribbon}}) / (L \cdot t_{\text{sheet}})$ (26) and assuming a 13-nm-thick nanoribbon, we obtain a sheet resistance of 65 kilohm, in good agreement with the measurements reported in Fig. 4A for microscopic squares of TCNL- rGO_{epi} .

TCNL does not require any solvents or lithographic resists that could contaminate the sample. This is especially important because the electronic properties of graphene vary strongly with surface doping. The strategy of variably reducing extended GO films is a general one that could be implemented in multiple ways depending on the application. The manufacture of graphene nanoelectronics could be achieved by

using arrays of heated probe tips (18). Independently addressed heated probe tips could alternately read or write nanostructures on a surface and in large arrays could address wafer-scale areas at high speed. A nano-embossing approach might also achieve local GO reduction, provided that the imprint template would offer nanometer-scale control of the reducing temperature field.

References and Notes

1. C. Berger *et al.*, *Science* **312**, 1191 (2006).
2. J. Kedzierski *et al.*, *IEEE Trans. Electron. Dev.* **55**, 2078 (2008).
3. D. A. Dikin *et al.*, *Nature* **448**, 457 (2007).
4. S. Park, R. S. Ruoff, *Nat. Nanotechnol.* **4**, 217 (2009).
5. G. Eda, C. Mattevi, H. Yamaguchi, H. Kim, M. Chhowalla, *J. Phys. Chem. C* **113**, 15768 (2009).
6. X. S. Wu *et al.*, *Phys. Rev. Lett.* **101**, 026801 (2008).
7. A. J. Epstein, in *Conductive Polymers and Plastics: In Industrial Applications*, L. Rupprecht, Ed. (William Andrew, Norwich, NY, 1999), pp. 1–4.
8. C. E. Parman, N. E. Israeloff, J. Kakalios, *Phys. Rev. B* **47**, 12578 (1993).
9. J. T. Robinson, F. K. Perkins, E. S. Snow, Z. Q. Wei, P. E. Sheehan, *Nano Lett.* **8**, 3137 (2008).
10. J. T. Robinson *et al.*, *Nano Lett.* **8**, 3441 (2008).
11. D. Wang *et al.*, *Appl. Phys. Lett.* **95**, 233108 (2009).
12. D. Wang *et al.*, *Adv. Funct. Mater.* **19**, 3696 (2009).
13. R. Szooskiewicz *et al.*, *Nano Lett.* **7**, 1064 (2007).
14. D. Wang *et al.*, *Appl. Phys. Lett.* **91**, 243104 (2007).
15. P. E. Sheehan, L. J. Whitman, W. P. King, B. A. Nelson, *Appl. Phys. Lett.* **85**, 1589 (2004).
16. M. Yang, P. E. Sheehan, W. P. King, L. J. Whitman, *J. Am. Chem. Soc.* **128**, 6774 (2006).
17. B. A. Nelson, W. P. King, A. R. Laracuente, P. E. Sheehan, L. J. Whitman, *Appl. Phys. Lett.* **88**, 033104 (2006).
18. P. Vettiger *et al.*, *IEEE Trans. NanoTechnol.* **1**, 39 (2002).
19. C. Gómez-Navarro *et al.*, *Nano Lett.* **7**, 3499 (2007).
20. C. Mattevi *et al.*, *Adv. Funct. Mater.* **19**, 2577 (2009).
21. See supporting material available on Science Online.
22. G. Eda, G. Fanchini, M. Chhowalla, *Nat. Nanotechnol.* **3**, 270 (2008).
23. H. C. Schniepp *et al.*, *J. Phys. Chem. B* **110**, 8535 (2006).
24. J. C. Meyer *et al.*, *Nature* **446**, 60 (2007).
25. T. Fillette *et al.*, *Phys. Rev. Lett.* **102**, 086102 (2009).
26. The thickness of the ribbon, t_{ribbon} , and the thickness of the sheet, t_{sheet} , have been inferred from the thickness of the corresponding GO films as reported in the SOM.
27. This work has been supported by the National Science Foundation (CMDITR program DMR 0120967, Materials Research Science and Engineering Center program DMR

0820382, and DMR-0706031), U.S. Department of Energy (DE-FG02-06ER46293 and PECASE), the Institute for Nanoscience at Naval Research Laboratory (NRL), the Office of Naval Research of the United States, the Defense Advanced Research Projects Agency (DARPA) Tip-Based Nanomanufacturing program, and Georgia Institute of Technology (Georgia Tech Research Foundation, COE Cutting Edge Research Award, and COPE

fellowship). This research was performed while Z.W. held a National Research Council Research Associateship Award at Naval Research Laboratory of the United States. Z.W. thanks H. Qi at NRL for providing doped silicon wafers, evaporating gold layers on mica, and taking scanning electron micrograph images of GO sheets on gold and silicon substrates. E.R. thanks T.-D. Li and M. Lucas for support with CAFM.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1373/DC1
Materials and Methods
Figs. S1 to S8
References and Notes

9 February 2010; accepted 29 April 2010
10.1126/science.1188119

Quaternary Ammonium (Hypo)iodite Catalysis for Enantioselective Oxidative Cycloetherification

Muhammet Uyanik,¹ Hiroaki Okamoto,¹ Takeshi Yasui,¹ Kazuaki Ishihara^{1,2*}

It is desirable to minimize the use of rare or toxic metals for oxidative reactions in the synthesis of pharmaceutical products. Hypervalent iodine compounds are environmentally benign alternatives, but their catalytic use, particularly for asymmetric transformations, has been quite limited. We report here an enantioselective oxidative cycloetherification of ketophenols to 2-acyl-2,3-dihydrobenzofuran derivatives, catalyzed by in situ-generated chiral quaternary ammonium (hypo)iodite salts, with hydrogen peroxide as an environmentally benign oxidant. The optically active 2-acyl-2,3-dihydrobenzofuran skeleton is a key structure in several biologically active compounds.

Over the past two decades, hypervalent iodine compounds have been increasingly explored as environmentally benign oxidation reagents in place of rare or toxic heavy metal oxidants (1, 2). However, their stoichiometric use has been limited because of potentially explosive shock-sensitivity and/or poor solubility in common organic solvents (1, 2). Thus, the development of hypervalent iodine-catalyzed reactions using more convenient stoichiometric co-oxidants is needed (3, 4). Harnessing chiral hypervalent iodine compounds for enantioselective oxidative coupling has proven a particular challenge in asymmetric catalysis. There are several examples of catalysis by in situ-generated chiral aryl- λ^3 - or aryl- λ^5 -iodane (5) with *meta*-chloroperbenzoic acid (*m*-CPBA) as a co-oxidant (Fig. 1A, left) (6–9); these include Quideau *et al.*'s enantioselective hydroxylative dearomatization of phenols (6), Altermann *et al.*'s enantioselective α -oxysulfonylation of ketones (7), and Dohi *et al.*'s and our independently reported enantioselective oxidative spirocyclizations of 1-naphthol derivatives (8, 9). In contrast, no strong examples have emerged of asymmetric catalysis using chiral cations paired with inorganic iodine-derived oxoacids, such as hypoiodous acid [IOH, I(I)], iodic acid [O=IOH, I(III)], iodic acid [(O=)₂IOH, I(V)], and periodic acid [(O=)₃IOH, I(VII)].

We report here an implementation of this strategy, using the atom-economical hydrogen peroxide as a mild stoichiometric oxidant to activate

catalytic ion pairs of chiral quaternary ammonium iodide (Fig. 1A, right) (10, 11). Specifically, we targeted enantioselective oxidative cycloetherification of ketophenols to 2-acyl-2,3-dihydrobenzofuran derivatives, using a *C*₂-symmetric chiral binaphthyl-based quaternary ammonium (hypo)iodite catalyst generated in situ by reaction with hydrogen

peroxide (Fig. 1B). The chiral 2-substituted 2,3-dihydrobenzofuran skeleton is a key structure in several biologically active compounds of medicinal interest (12–19). Earlier preparations of optically active 2-alkenyl-2,3-dihydrobenzofuran derivatives have relied on transition metal catalysis (20–23).

The catalytic or stoichiometric oxidation of 3-(2-hydroxyphenyl)-1-phenylpropan-1-one (1) with phenyl- λ^3 -iodanes gave a complex mixture, and the desired (2,3-dihydrobenzofuran-2-yl) (phenyl)methanone (2) was not detected (Fig. 2A, entries 1 and 2). In sharp contrast, and to our delight, the oxidation of 1 with two equivalents of hydrogen peroxide [30 weight percent (wt %) in water] in the presence of 10 mole percent (mol %) of tetrabutylammonium iodide (Bu₄NI) in acetonitrile at room temperature gave 2 in 87% yield (Fig. 2A, entry 3) (24). Furthermore, the oxidation of 1 was much faster in diethyl ether (Et₂O), tetrahydrofuran (THF), or ethyl acetate (EtOAc) (Fig. 2A, entry 4, and table S1). Notably, the oxidation of 1 did not occur on substitution of tetrabutylammonium bromide or chloride for Bu₄NI. Excellent chemoselectivity was observed

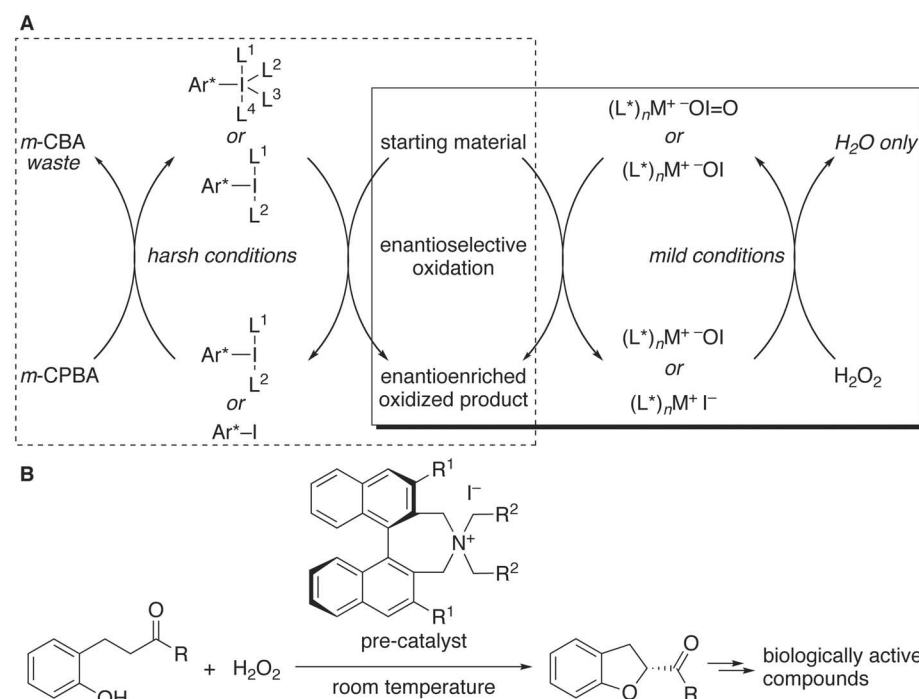
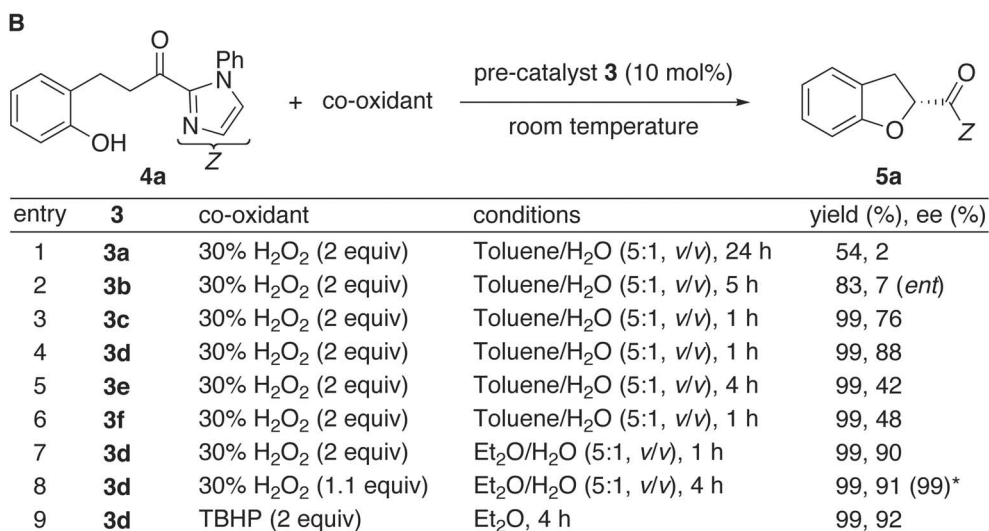
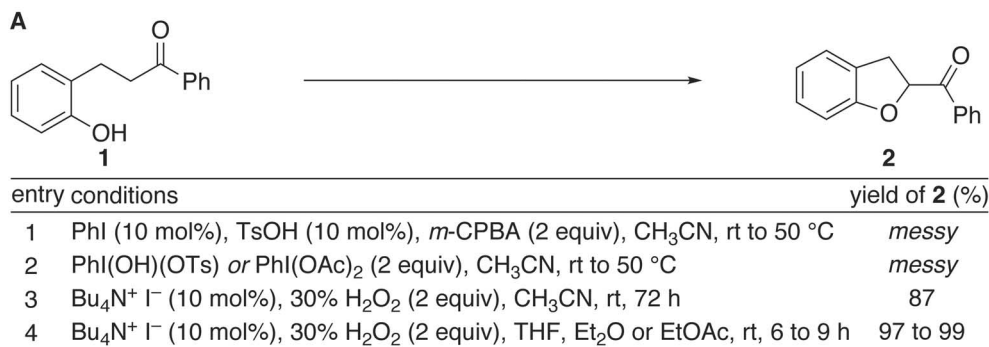


Fig. 1. (A) (Left) Known in situ-generated aryl- λ^3 - or aryl- λ^5 -iodane catalysis. (Right) In situ-generated hypoiodite(I) or iodite(III) catalysis. (B) Design of inorganic iodide precatalyst paired with a chiral quaternary ammonium counterion for the enantioselective oxidative cycloetherification of ketophenols to 2-acyl-2,3-dihydrobenzofuran derivatives. Ar, aryl; L, ligand; M⁺, metal or onium cation, R, alkyl or aryl group. Symbols (Ar and L) marked with asterisks represent chiral groups.

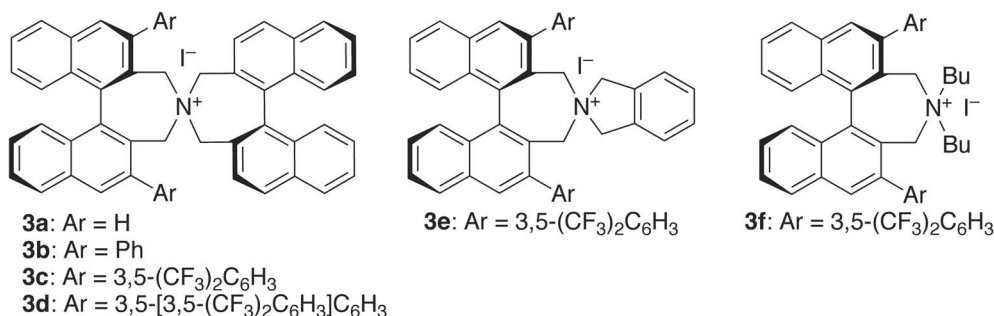
¹Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa, Nagoya, 464-8603, Japan. ²Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Furo-cho, Chikusa, Nagoya, 463-8603, Japan.

*To whom correspondence should be addressed. E-mail: ishihara@cc.nagoya-u.ac.jp

Fig. 2. (A) Oxidative cycloetherification of **1**. Isolated yields of **2** are reported. Ts, *p*-toluenesulfonyl. **(B)** Enantioselective oxidative cycloetherification of **4a**. Isolated yields of **2** are reported. ee was determined by chiral stationary-phase high-performance liquid chromatography (HPLC). Ph, phenyl; Pr, propyl; Bu, butyl; h, hours.



* After a single recrystallization from *i*-PrOH/EtOH (4:1, v/v).



under the present oxidation. We detected no phenol oxidation products, which are well known to form in reactions with aryl-λ³-iodanes (1–9).

To render the catalysis asymmetric, we examined N-spiro-type quaternary ammonium iodide precatalysts (**3**) bearing chiral 3,3'-disubstituted 1,1'-binaphthyl skeletons, analogous to Ooi and Maruoka's chiral phase-transfer catalysts (Fig. 2B) (25). As a result, we found that a 1-phenyl-1*H*-imidazol-2-yl moiety (*Z*) at the 1-position of the substrates (**4a**) was effective for inducing high enantioselectivity (26) (Fig. 2B and table S3). The oxidation of **4a** with two equivalents of hydrogen peroxide (30 wt % in water) in the presence of 10 mol % of **3a** in mixed toluene/water solvent (5/1, v/v) gave **2a** in 54% yield, but with very low enantioselectivity [2% enantiomeric excess (ee)] (Fig. 2B, entry 1). The substituents at the 3,3'-positions of the binaphthyl moiety were

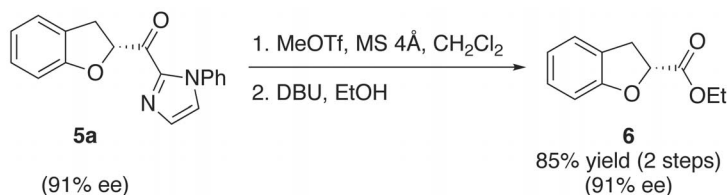


Fig. 3. Conversion of **5a** to (*R*)-ethyl ester **6**. MeOTf, methyl trifluoromethanesulfonate; MS, molecular sieves; DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene.

crucial for increasing not only the enantioselectivity but also the chemical yield of **5a** (Fig. 2B, entries 1 to 4). Ammonium cations bearing bulky and electron-deficient substituents {Ar = 3,5-[3,5-(CF₃)₂C₆H₃]₂C₆H₃} at the 3,3'-positions gave the best results (1 hour, 99% yield, 88% ee) (Fig. 2B, entry 4). The use of mono(1,1'-binaphthyl-2,2'-dimethyl)ammonium iodides such as **3e** and

3f in place of bis(1,1'-binaphthyl-2,2'-dimethyl) ammonium iodide (**3c**) gave moderate enantioselectivities (Fig. 2B, entries 5 and 6 versus entry 3). Notably, moderate-to-high enantioselectivities were observed, regardless of the polarity of the solvent (table S3). In particular, **5a** was obtained in 99% yield with 90% ee in mixed diethyl ether/water solvent (5/1, v/v) (Fig. 2B, entry 7). 1.1 equiv-

alent of hydrogen peroxide was suitable as a co-oxidant for quantitative conversion (Fig. 2B, entry 8). Importantly, nearly enantiomerically pure **5a** was obtained after a single recrystallization (99% ee) (Fig. 2B, entry 8). When anhydrous *tert*-butylhydroperoxide (TBHP) was used as a co-oxidant in place of aqueous hydrogen perox-

ide, **5a** was obtained with higher enantioselectivity (92% ee) in diethyl ether (Fig. 2B, entry 9).

The **5** products are very useful chiral intermediates for further synthetic elaboration (26). For instance, **5a** was transformed efficiently into the known (*R*)-ethyl ester (**6**) (12, 27), which is a synthetic intermediate for natural products such

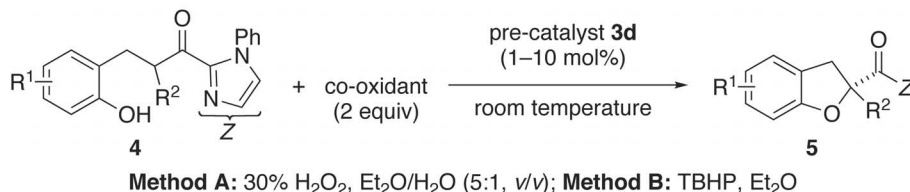
as tremetone (12, 22), fomannoxin (28), and anodendroic acid (Fig. 3) (28).

To explore the generality and scope of the present enantioselective oxidative cyclization, we prepared and examined several ketophenol derivatives (**4**) as substrates under optimized conditions (Table 1). The oxidation of 4-substituted phenol derivatives **4b** to **4f** bearing electron-donating or electron-withdrawing groups gave the corresponding dihydrofuran derivatives **5b** to **5f** in excellent yields with high enantioselectivities (91 to 93% ee) (Table 1, entries 1 to 7). In most cases, we observed high enantioselectivities using method A or B, though method B offered substantially higher yield and ee in the case of the 3,5-dimethoxy-substituted phenol derivative **4g** (Table 1, entries 8 and 9). The 2-alkenyl-derivative of compound **5g** is a synthetic intermediate for natural products such as *ent*-remirol and *ent*-remiridol (29). The reaction was effective for the 2-naphthol derivative **4i** (Table 1, entry 11), as well as for penta-substituted phenol derivative **4j** (Table 1, entry 12). Oxidation of 2-methyl-substituted substrate **4k** gave a product (**5k**) bearing a tetrasubstituted stereogenic center in 96% ee (Table 1, entry 13). The precatalyst loading of **3d** could be reduced to 1 mol % without affecting the chemical yield and enantioselectivity (Table 1, entry 14). Compound **5k** and its analogs would potentially offer a different route to biologically active targets, such as peroxisome proliferator-activated receptor α agonists (15).

To understand the reaction pathway, we conducted several control experiments directed toward identifying an active hypervalent iodine species (table S4). No oxidation of **1** occurred in the presence of stoichiometric amounts of tetrabutylammonium iodate(V) or periodate(VII). Thus, the iodate(V) and periodate(VII) active species were ruled out. Additionally, the stoichiometric oxidation of **1** with *N*-iodosuccinimide or molecular iodine (I_2) gave a complex mixture, and desired product **2** was not detected. Notably, no iodinated products were detected, even in the oxidation of electron-rich phenol **4h** under our catalytic conditions (Table 1, entries 8 and 9). In contrast, the oxidation of **1** with I_2 in the presence of two equivalents of tetrabutylammonium hydroxide ($[Bu_4N]^+[OH]^-$) gave desired product **2** in 91% yield (30, 31). This result suggests that in situ-generated tetrabutylammonium hypoiodite $\{[Bu_4N]^+[IO]^-$, $I(I)\}$ may be an active oxidant species for **1**. However, $[Bu_4N]^+[IO]^-$ may also disproportionate to tetrabutylammonium iodite $\{[Bu_4N]^+[IO_2]^-$, $I(III)\}$ and Bu_4NI under our reaction conditions, implying that $[Bu_4N]^+[IO_2]^-$ is another potential actual oxidant species. Thus, we propose a catalytic cycle involving chiral quaternary ammonium hypoiodite ($[R_4N]^+[IO]^-$) or iodite ($[R_4N]^+[IO_2]^-$), which should be generated in situ from ammonium iodide (R_4NI) and a co-oxidant (Fig. 1A, right).

These results highlight the substantial scope of chiral salt catalysis using inorganic iodine-derived oxoacids in place of aryl iodane or transition metal catalysts.

Table 1. Scope of the enantioselective oxidative cycloetherification of **4**. Unless otherwise noted, the reactions were performed with 0.1 mmol of **4** and 0.2 mmol of co-oxidant in the presence of **3d** (10 mol %). The isolated yields of **5** are reported. ee was determined by chiral stationary-phase HPLC. Bn, benzyl.



Entry	Product	5	Method, time (h)	Yield (%)	ee (%)
1		5b	A, 2	99	93
2		5c	A, 2	99	91
3		5d	A, 2	99	91
4		5e	A, 1.5	84	85
5		5e	B, 4	99	91
6		5f	A, 2	98	89
7		5f	B, 6.5	99	92
8		5g	A, 2	26	70
9		5g	B*, 0.5	99	85
10		5h	A, 2	99	90
11		5i	A, 2.5	99	89
12		5j	A, 1.5	78	87
13		5k †	A, 3	99	96
14†		5k †	B, 17	99	96

*Reaction was performed in toluene instead of Et_2O .

†The reaction was performed with 2 mmol of **5k** with 4 mmol of TBHP in the presence of 1 mol % of **3d**.

‡Absolute configuration of **5k** was determined by transformation into the known (*R*)-methyl ketone (24).

References and Notes

1. T. Wirth, Ed., *Hypervalent Iodine Chemistry (Topics in Current Chemistry)* (Springer, Berlin, 2003), vol. 224.
2. V. V. Zhdankin, P. J. Stang, *Chem. Rev.* **108**, 5299 (2008), and references therein.
3. M. Ochiai, K. Miyamoto, *Eur. J. Org. Chem.* **2008**, 4229 (2008), and references therein.
4. T. Dohi, Y. Kita, *Chem. Commun. (Cambridge)* **2009**, 2073 (2009), and references therein.
5. According to International Union of Pure and Applied Chemistry rules, compounds with nonstandard bonding numbers are named using λ notation; thus, tri- and penta-valent iodine compounds are designated λ^3 - and λ^5 -iodanes, indicating that the iodine atoms have 10 and 12 valence electrons, respectively.
6. S. Quideau *et al.*, *Angew. Chem. Int. Ed.* **48**, 4605 (2009).
7. S. M. Altermann *et al.*, *Eur. J. Org. Chem.* **2008**, 5315 (2008).
8. T. Dohi *et al.*, *Angew. Chem. Int. Ed.* **47**, 3787 (2008).
9. M. Uyanik, T. Yasui, K. Ishihara, *Angew. Chem. Int. Ed.* **49**, 2175 (2010).
10. For the oxidation of iodide or iodine with hydrogen peroxide, see (11).
11. I. Matsuzaki, T. Nakajima, H. A. Liebhafsky, *Chem. Lett.* **3**, 1463 (1974), and references therein.
12. D. M. Bowen, J. I. DeGraw Jr., V. R. Shah, W. A. Bonner, *J. Med. Chem.* **6**, 315 (1963).
13. A. H. Banskota *et al.*, *J. Nat. Prod.* **61**, 896 (1998).
14. C. L. Céspedes, A. Uchoa, J. R. Salazar, F. Perich, F. Pardo, *J. Agric. Food Chem.* **50**, 2283 (2002).
15. G. Q. Shi *et al.*, *J. Med. Chem.* **48**, 5589 (2005).
16. G.-H. Chu *et al.*, *Bioorg. Med. Chem. Lett.* **15**, 5114 (2005).
17. J. L. Cohen, A. Limon, R. Miledi, A. R. Chamberlin, *Bioorg. Med. Chem. Lett.* **16**, 2189 (2006).
18. C. Ito *et al.*, *J. Nat. Prod.* **69**, 138 (2006).
19. J. A. Duan, L. Wang, S. Qian, S. Su, Y. Tang, *Arch. Pharm. Res.* **31**, 965 (2008).
20. Y. Uozumi, K. Kato, T. Hayashi, *J. Am. Chem. Soc.* **119**, 5063 (1997).
21. R. M. Trend, Y. K. Ramtohl, E. M. Ferreira, B. M. Stoltz, *Angew. Chem. Int. Ed.* **42**, 2892 (2003).
22. S. C. Pelly, S. Govender, M. A. Fernandes, H.-G. Schmalz, C. B. de Koning, *J. Org. Chem.* **72**, 2857 (2007).
23. S. Tanaka, T. Seki, M. Kitamura, *Angew. Chem. Int. Ed.* **48**, 8948 (2009), and references therein.
24. Materials and methods are available as supporting material on Science Online.
25. T. Ooi, K. Maruoka, *Angew. Chem. Int. Ed.* **46**, 4222 (2007), and references therein.
26. D. A. Evans, K. R. Fandrick, H.-J. Song, K. A. Scheidt, R. Xu, *J. Am. Chem. Soc.* **129**, 10029 (2007).
27. W. A. Bonner *et al.*, *Tetrahedron* **20**, 1419 (1964).
28. Y. Kawase, S. Yamaguchi, O. Inoue, M. Sannomiya, K. Kawabe, *Chem. Lett.* **9**, 1581 (1980).
29. S. Yamaguchi, S. Muro, M. Kobayashi, M. Miyazawa, Y. Hirai, *J. Org. Chem.* **68**, 6274 (2003).
30. Potassium hypoiodite (IOK) is generated in situ from I_2 and two equivalents of KOH, see (31).
31. S. Yamada, D. Morizono, K. Yamamoto, *Tetrahedron Lett.* **33**, 4329 (1992).
32. Financial support for this project was partially provided by a Grant-in-Aid for Scientific Research (#20245022) from the Japan Society for the Promotion of Science, the New Energy and Industrial Technology Development Organization, the Shionogi Award in Synthetic Organic Chemistry, Japan, and the Global Centers of Excellence Program in Chemistry of Nagoya University from the Ministry of Education, Culture, Sports, Science and Technology.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1376/DC1
Materials and Methods

SOM Text

Scheme S1

Tables S1 to S4

References

11 February 2010; accepted 23 April 2010
10.1126/science.1188217

Regulation of Body Temperature by Some Mesozoic Marine Reptiles

Aurélien Bernard,¹ Christophe Lécuyer,^{1,2*} Peggy Vincent,³ Romain Amiot,¹ Nathalie Bardet,³ Eric Buffetaut,⁴ Gilles Cuny,⁵ François Fourel,¹ François Martineau,¹ Jean-Michel Mazin,¹ Abel Prieur¹

What the body temperature and thermoregulation processes of extinct vertebrates were are central questions for understanding their ecology and evolution. The thermophysiology status of the great marine reptiles is still unknown, even though some studies have suggested that thermoregulation may have contributed to their exceptional evolutionary success as apex predators of Mesozoic aquatic ecosystems. We tested the thermal status of ichthyosaurs, plesiosaurs, and mosasaurs by comparing the oxygen isotope compositions of their tooth phosphate to those of coexisting fish. Data distribution reveals that these large marine reptiles were able to maintain a constant and high body temperature in oceanic environments ranging from tropical to cold temperate. Their estimated body temperatures, in the range from $35^\circ \pm 2^\circ\text{C}$ to $39^\circ \pm 2^\circ\text{C}$, suggest high metabolic rates required for predation and fast swimming over large distances offshore.

The metabolic status of extinct vertebrates is a key to understand their feeding strategy, which was critical for satisfying their daily energy requirements, as well as their potential to exploit cold environments. Phylogeny and ecology most likely had a large influence on the thermophysiology of past vertebrates. High metabolic rates mean the need to access large amounts of high-quality food, which may be sat-

isfied by adopting predatory behavior, as shared by many carnivorous mammals, except scavengers. Endothermy is the ability to generate and retain enough heat to elevate body temperature to a high but stable level, whereas homeothermy is the maintenance of a constant body temperature in different thermal environments (1, 2). Such internal production of heat is not restricted to mammals and birds. Heat generation can have several origins: digestive organs in mammals and birds (1) or muscles in endothermic lamniform sharks (3). Paladino *et al.* (4) proposed that some marine reptiles such as leatherback turtles display endothermy instead of inertial homeothermy, thus helping them to feed in cold waters. However, Lutcavage *et al.* (5) showed that the studied gravid female specimens raised their metabolic rates because of egg laying, thus biasing the evaluation of their true metabolic status. Most biologists agree that full or incomplete en-

dothermy arose several times during species evolution and developed independently in several lineages. For example, partial endothermy is known in sharks, tunas, and even in some insects and flowers (6–8). The origin and spreading of endothermy are still a matter of great debate (9–11); its oldest occurrence could be as early as the Permian, with the appearance and radiation of the Synapsida. Among archosaurs, mass homeothermy or even endothermy have been proposed for dinosaurs (12) and pterosaurs (13) and suggested for the ancestors of crocodylians, because of the existence of a four-chambered heart, which modern crocodyles share with mammals and birds (11).

Large marine reptiles, including ichthyosaurs, plesiosaurs, and mosasaurs, inhabited the oceans from the Triassic to the Cretaceous. They represent three different lineages that became secondarily adapted to a marine mode of life. Ichthyosaurs evolved from basal neodiapsid reptiles, with the most obvious aquatic adaptations: a dolphin-like streamlined body without a neck, paddles, and a fish-like tail. Plesiosaurs are derived diapsids, which belong to the Sauropterygia, the sister group of the Lepidosauria (lizards and snakes). They are highly adapted for submarine locomotion, with powerful paddle-like limbs and heavily reinforced limb girdles. Motani (14) already discussed the possibility that plesiosaurs could not have had a typical reptilian physiology, thus indicating high metabolic activity. Mosasaurs constitute a family of Late Cretaceous varanoid anguimorphs highly adapted to marine life; they are derived lepidosaurs with an elongate body, deep tail, and paddle-like limbs (15). Both tooth morphology and the stomach contents of these three groups of marine reptiles indicate predatory behavior. Their anatomy could afford high cruising speeds and a basal metabolic rate similar to that of modern tunas (16). Moreover, the bone structure of adult

¹CNRS UMR 5125, Paléoenvironnements et Paléobiosphère, Université Lyon 1, Lyon, 2 Rue Raphaël Dubois, 69622 Villeurbanne Cedex, France. ²Institut Universitaire de France, 103 Boulevard Saint-Michel, 75005 Paris, France. ³CNRS UMR 5143, Département Histoire de la Terre, Muséum National d'Histoire Naturelle, 8 Rue Buffon 75231 Paris, France. ⁴CNRS UMR 8538, Laboratoire de Géologie de l'Ecole Normale Supérieure, 24 Rue Lhomond, 75231 Paris, France. ⁵Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark.

*To whom correspondence should be addressed. E-mail: clecuyer@univ-lyon1.fr

plesiosaurs and mosasaurs corresponds to that of large pelagic marine predators designed for long cruises in open waters (17). The stomach contents of these marine reptiles have revealed that they were predators feeding on highly diverse foods, including other marine reptiles, fish, cephalopods, and crinoids (18). The metabolic status of these aquatic reptiles was also investigated through histological bone studies. It has been shown that Jurassic ichthyosaurs had rapid post-natal growth, followed by intense bone remodeling that could be related to a sustained metabolic rate close to that of marine endotherms (9). Adaptation to cold marine waters was also revealed by the fossil reptile assemblage discovered in the Aptian southern high-latitude deposits of the White Cliffs in southeast Australia (19). The specimens were attributed to at least three families of plesiosaurs and at least one of ichthyosaurs. Paleoclimatic proxies indicate cold to near-freezing conditions at the seasonal scale, a climate mode that is not tolerated by modern ectothermic reptiles such as turtles or crocodiles. This observation suggests that some Mesozoic marine reptile taxa were able to cope with low-temperature marine environments (19). In this study, we investigated the metabolic status of ichthyosaurs, plesiosaurs, and mosasaurs using the oxygen isotope compositions ($\delta^{18}\text{O}$) of their phosphatic tissues.

The $\delta^{18}\text{O}$ value of vertebrate phosphate depends on both body temperature and the com-

position of ingested water (20). In the case of the studied reptiles, the estimated body temperatures recorded in the $\delta^{18}\text{O}$ value reflect that of the blood during tooth development; this should be also the case for some plesiosaurs that had protruding teeth. Within each studied reptile group, the

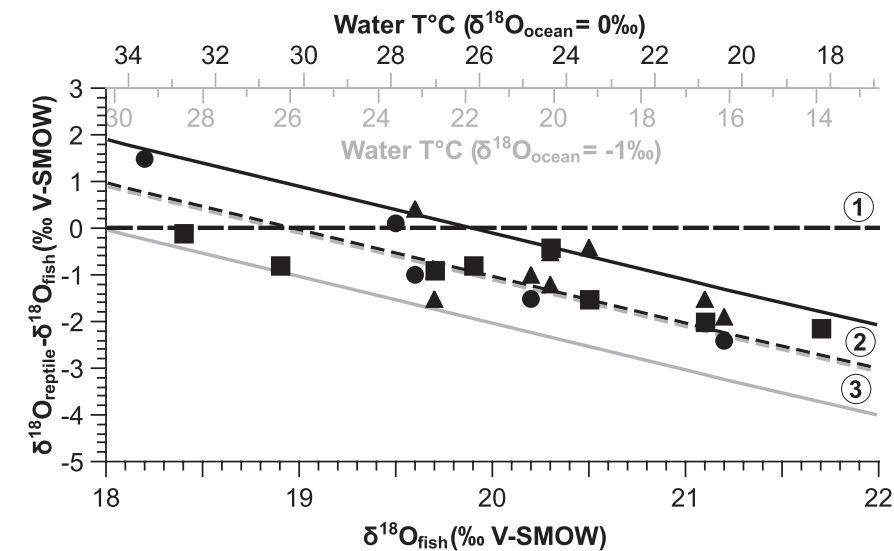


Fig. 1. Model variation of the differences in the $\delta^{18}\text{O}$ of tooth phosphate between marine reptiles and fish against the variation of the $\delta^{18}\text{O}$ of fish teeth, assuming (1) an ectothermic and poikilothermic reptile [body water $\delta^{18}\text{O}$ and body temperature (T) equal seawater $\delta^{18}\text{O}$ and seawater temperature]; (2) an endothermic reptile with body temperature ranging from 35°C (solid black line) to 39°C (dashed black line) and body water 2‰ enriched relative to a seawater value of 0‰; and (3) an endothermic reptile with body temperature ranging from 35°C (dashed gray line) to 39°C (solid gray line) and body water 2‰ enriched relative to a seawater value of -1‰. For comparison, ichthyosaur (circles), plesiosaur (triangles), and mosasaur (squares) values are reported. V-SMOW, Vienna SMOW values.

Table 1. Mean $\delta^{18}\text{O}$ values of tooth phosphate from worldwide Mesozoic ichthyosaurs, mosasaurs, and plesiosaurs, as well as coexisting marine fish.

Locality number	Locality	Age	Fish $\delta^{18}\text{O}$			Ichthyosaur $\delta^{18}\text{O}$			Plesiosaur $\delta^{18}\text{O}$			Mosasaur $\delta^{18}\text{O}$			Reference
			<i>n</i>	$\delta^{18}\text{O}$	SD	<i>n</i>	$\delta^{18}\text{O}$	SD	<i>n</i>	$\delta^{18}\text{O}$	SD	<i>n</i>	$\delta^{18}\text{O}$	SD	
1	South Dakota, USA	Early Campanian	2	19.8	0.1							1	18.5		(24)
2	Kansas, USA	Late Coniacian	3	18.9	0.8							1	18.1		(24)
3	Oulad Abdoun, Morocco	Maastrichtian	5	20.3	0.9				3	19.8	0.6	1	19.9		This study
4	Cambridge, UK	Late Albian	2	20.2	0.1	3	18.7	0.2	2	19.2	0.6				This study
5	Kimmeridge Clay, Westbury, UK	Kimmeridgian	4	19.5	0.1	2	19.6	0.6							This study
6	Oxford Clay, Peterborough, UK	Early Callovian	5	19.6	0.3	4	18.6	1.4	2	20.0	0.4				(23)
7	Sorel, France	Sinemurian	1	19.7					1	18.8					This study
8	Crussol, France	Middle Oxfordian	2	20.3	1.0				1	19.1					This study; (26)
9	Bourgogne, France	Early Oxfordian	1	21.2		1	18.8		1	19.3					This study
10	Maastricht, Netherlands	Maastrichtian	1	21.7					2	19.6					This study
11	Monte San Giorgio, Switzerland	Anisian	1	18.2		2	19.7	0.4							(27)
12	Asen, Sweden	Campanian	4	21.1	0.9				2	19.6	0.4	2	19.1	0.1	This study
13	Ullstorp, Sweden	Campanian	4	20.5	0.6				2	20.1	0.5	3	19.0	0.1	This study
14	Zefa, Israel	Late Campanian	3	19.7	0.4							1	18.8		(25)
15	Ruseifa, Jordan	Late Campanian	1	18.4								1	18.3		(25)
16	Toolebuc Formation, Australia	Albian	1	19.7					2	18.2	0.1				This study

taxonomic resolution is at the family level. Assuming that both reptiles and fish lived in the same water mass, differences in their $\delta^{18}\text{O}$ values would reflect differences in body temperature. To estimate the dependence of reptile body temperature on that of ambient water, we reported the difference in $\delta^{18}\text{O}$ value between coexisting marine

reptiles and fish (belonging to the same sedimentary bed) as a function of the fish $\delta^{18}\text{O}$ value, which has been proven to be a valuable proxy of seawater temperature (21). A compilation was performed by combining 53 new (22) and 27 published (23–27) (table S1) $\delta^{18}\text{O}$ values of coexisting marine fish and reptile tooth remains

recovered from worldwide sedimentary deposits of Triassic to Cretaceous ages. Figure 1 illustrates the principles of the relationship between coexisting reptile and fish temperature differences and seawater temperature. If reptile body temperatures mimic those of fish, $\delta^{18}\text{O}$ pairs are expected to plot along the horizontal null line. In the hypothetical scenario where reptile body temperature is constant and nearly independent of ambient water temperature, the isotopic pairs should lie on or close to a line with a slope of -1 . Paired $\delta^{18}\text{O}$ data (Table 1) are compatible with linear distributions whose slopes for ichthyosaurs, plesiosaurs, and mosasaurs are -1.38 ± 0.20 [coefficient of determination (R^2) = 0.937], -1.08 ± 0.27 (R^2 = 0.480), and -0.70 ± 0.20 (R^2 = 0.588), respectively. Significant scattering in paired data is observed in Fig. 2, and it exceeds uncertainties associated with analytical measurements. Diagenetic alteration cannot be excluded for some samples, even though tooth enamel was favored because of its remarkable resistance to postdepositional alteration and recrystallization (28). Reptiles and fish collected from the same sedimentary bed may not be strictly contemporaneous, depending on how much time was condensed in the sedimentary layer. Contemporaneous reptiles and fish could have recorded distinct sea surface temperatures or water $\delta^{18}\text{O}$, because they lived at various depths or migrated seasonally for hunting or reproduction. However, the observed linear correlations are robust enough to indicate that reptile body temperature does not vary significantly with seawater temperature, except for mosasaurs, whose slope could suggest that body temperature could slightly decrease with decreasing ambient water temperature. Indeed, the range in $\delta^{18}\text{O}$ of fish close to 3 per mil (‰) means a temperature variation of 13°C, according to the slope of the oxygen isotope fractionation equation between fish phosphate and water [temperature (°C) = $113.3 - 4.38 (\delta^{18}\text{O}_{\text{phosphate}} - \delta^{18}\text{O}_{\text{water}})$] that was determined by Kolodny *et al.* (21). This temperature range is valid only if the $\delta^{18}\text{O}$ value of surface marine waters was constant at various latitudes and for distinct water masses.

It is known that budgets of evaporation and precipitation over the oceans are responsible for various trends between $\delta^{18}\text{O}$ values of seawater and salinity (29). These relationships are difficult to apply to the past, especially for geological periods as old as the Mesozoic. At first order, we can consider that there is an ^{18}O enrichment of water by at least 1‰ relative to the mean ocean composition under low latitudes. Seawater tends to be ^{18}O -depleted by at least 1‰ when reaching latitudes above 50° as a consequence of precipitation dominating over evaporation; the $\delta^{18}\text{O}$ value can be even lower in the presence of continental masses with fluvial discharge, as observed in the present-day North Atlantic ocean off Canada, Greenland, and Norway. Consequently, the range in Mesozoic fish $\delta^{18}\text{O}$ values must be corrected by at least 2‰ because of the $\delta^{18}\text{O}$ -salinity latitudinal gradient. In other words, the observed

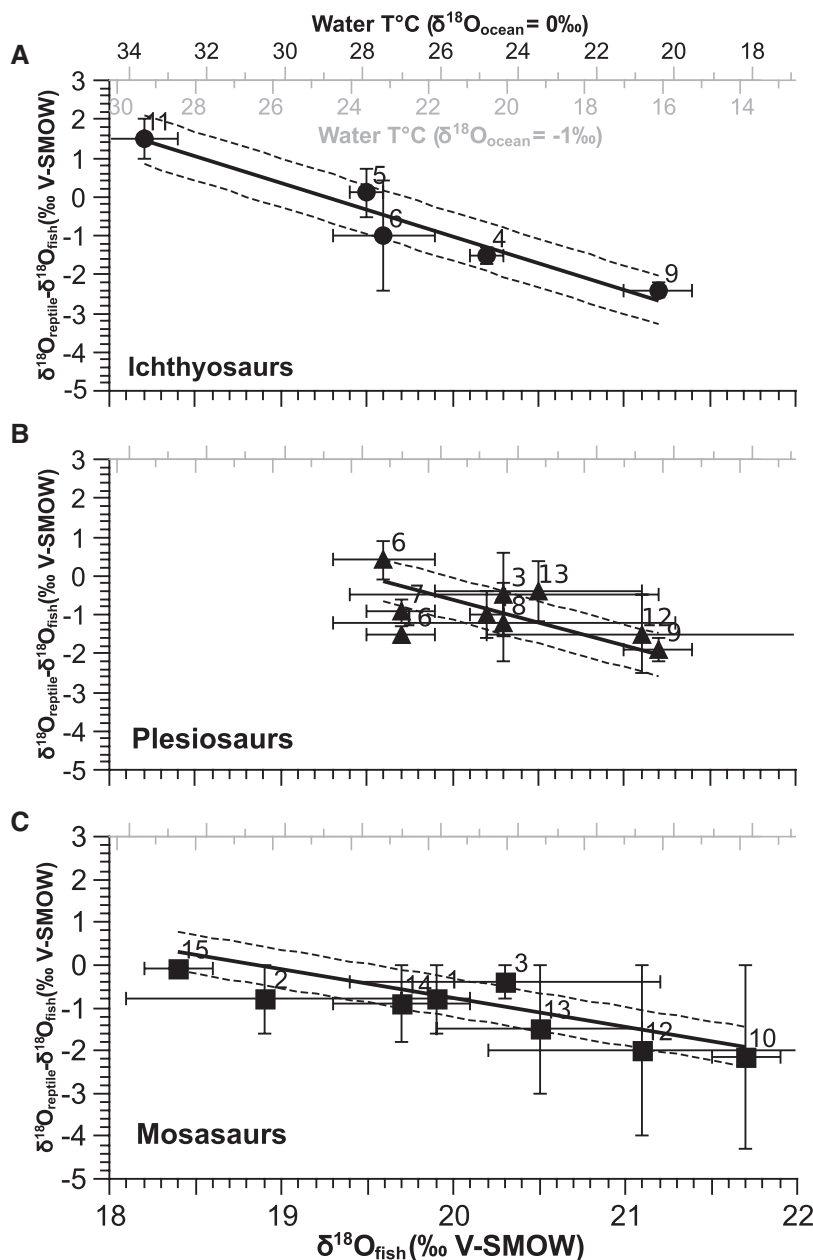


Fig. 2. Differences in the $\delta^{18}\text{O}$ of tooth phosphate between the three marine reptiles [(A) ichthyosaurs, (B) plesiosaurs, and (C) mosasaurs] and fish are reported against the $\delta^{18}\text{O}$ of fish teeth, approximating the body temperature differences between coexisting reptiles and fish and the seawater temperature where they lived (upper axis). The following reduced major axis regression lines with their 95% confidence limits are drawn: $y = -1.38 (\pm 0.20) \times x + 26.53 (\pm 3.07)$, $R^2 = 0.94$ (ichthyosaurs); $y = -1.08 (\pm 0.27) \times x + 20.92 (\pm 5.02)$, $R^2 = 0.48$ (plesiosaurs); and $y = -0.70 (\pm 0.20) \times x + 12.89 (\pm 3.37)$, $R^2 = 0.59$ (mosasaurs). Numbers refer to localities given in Table 1. Ichthyosaur and fish samples 11 from Monte San Giorgio should be considered cautiously for the following reasons: (i) a poor knowledge of the water salinity and consequently of the $\delta^{18}\text{O}$ value of ambient water. (ii) Triassic ichthyosaurs were not tuna-shaped yet, so their thermoregulation was most likely not well developed and difficult to compare with that of other Jurassic specimens. However, removing these data does not significantly change the slope value of the regression line (-1.336 instead of -1.377).

range of fish $\delta^{18}\text{O}$ values is translated into a temperature range of at least 25°C . Lécuyer *et al.* (26) have shown that the $\delta^{18}\text{O}$ of the global ocean most likely ranged from -1 to 0‰ [standard mean ocean water (SMOW) values] throughout the Jurassic and Cretaceous. Accordingly, the lowest temperatures of $12^\circ \pm 2^\circ\text{C}$ correspond to the highest fish $\delta^{18}\text{O}$ values approaching 22‰ , and the highest temperatures of $36^\circ \pm 2^\circ\text{C}$ correspond to the lowest fish $\delta^{18}\text{O}$ values close to 18‰ (Fig. 2). Ichthyosaurs and plesiosaurs have $\delta^{18}\text{O}$ values similar to those of fish at a corresponding temperature range of $26^\circ \pm 2^\circ\text{C}$ estimated from fish and seawater $\delta^{18}\text{O}$ values of 19.5‰ and -1 to 0‰ , respectively (Fig. 2). The regression line for mosasaurs intercepts the horizontal null line at a lower $\delta^{18}\text{O}$ value of 18.5‰ , thus indicating a possible higher body temperature of $30^\circ \pm 2^\circ\text{C}$. These temperature ranges are the minimal values that can be considered for ichthyosaur, plesiosaur, and mosasaur body temperatures if the $\delta^{18}\text{O}$ of their body equaled that of ambient seawater.

However, aquatic breathing vertebrates have body waters that are slightly ^{18}O -enriched relative to ambient water in the absence of transcutaneous evapotranspiration, and published data for modern aquatic reptiles reveal that this isotopic enrichment does not exceed 2‰ (30, 31). Consequently, the body temperatures of studied ichthyosaurs and plesiosaurs could have been as high as $35^\circ \pm 2^\circ\text{C}$ and even close to $39^\circ \pm 2^\circ\text{C}$ for mosasaurs, according to Kolodny *et al.*'s equation (21). Both slope values of linear regressions and estimates of body temperatures are in good agreement with the swimming performances that were modeled for these three groups of marine reptiles. Massare (32, 33) and Motani (14) suggested that ichthyosaurs were pursuit predators, whereas most mosasaurs were ambush predators, not requiring high metabolic rates all the time. Plesiosaurs were considered to have been cruisers, although slower than ichthyosaurs in sustained speed.

The $\delta^{18}\text{O}$ values of Mesozoic ichthyosaurs and plesiosaurs support the hypothesis that these large predators were able to regulate their body temperature independently of the surrounding water temperature even when it was as low as about $12^\circ \pm 2^\circ\text{C}$. In the case of mosasaurs, we cannot exclude the possibility that their body temperature was partly influenced by the temperature of ambient water. In any case, estimated body temperatures in the range from $35^\circ \pm 2^\circ\text{C}$ to $39^\circ \pm 2^\circ\text{C}$ encompass those of modern cetaceans (34) and suggest a high metabolic rate required for predation and fast swimming over large distances, especially in cold waters. $\delta^{18}\text{O}$ data from tooth phosphate reveal the existence of homeothermy for ichthyosaurs and plesiosaurs, and of at least partial homeothermy for mosasaurs, with in all cases a taxonomic resolution that does not exceed the family or infraorder. These three distinct phylogenetic groups of large marine reptiles were able to maintain a body temperature substantially higher than that of ambient marine waters,

indicating that some kind of endothermy operated as an internal source of heat.

References and Notes

1. A. F. Bennett, J. W. Hicks, A. J. Cullum, *Evolution* **54**, 1768 (2000).
2. C. P. Hickman, L. S. Roberts, A. Larson, *Integrated Principles of Zoology* (McGraw-Hill Science, New York, 2000).
3. J. K. Carlson, K. J. Goldman, C. G. Lowe, in *Biology of Sharks and Their Relatives*, J. C. Carrier, J. A. Musick, M. R. Heithaus, Eds. (CRC Press, Boca Raton, FL, 2004), pp. 203–224.
4. F. V. Paladino, M. P. O'Connor, J. R. Spotila, *Nature* **344**, 858 (1990).
5. M. E. Lutcavage, P. G. Bushnell, D. R. Jones, *Can. J. Zool.* **70**, 348 (1992).
6. B. Heinrich, *Integr. Comp. Biol.* **29**, 1157 (1989).
7. B. A. Block, J. R. Finnerty, A. F. Stewart, J. Kidd, *Science* **260**, 210 (1993).
8. R. S. Seymour, *Plant Cell Environ.* **27**, 1014 (2004).
9. V. De Buffrénil, J. Mazin, *Paleobiology* **16**, 435 (1990).
10. R. S. Seymour, *Nature* **262**, 207 (1976).
11. R. S. Seymour, C. L. Bennett-Stamper, S. D. Johnston, D. R. Carrier, G. C. Grigg, *Physiol. Biochem. Zool.* **77**, 1051 (2004).
12. R. T. Bakker, *Nature* **238**, 81 (1972).
13. P. Wellnhofer, *The Illustrated Encyclopedia of Pterosaurs* (Crescent Books, New York, 1991).
14. R. Motani, *Paleobiology* **28**, 251 (2002).
15. M. J. Benton, *Vertebrate Palaeontology* (Wiley-Blackwell, Oxford, 2005).
16. R. Motani, *Nature* **415**, 309 (2002).
17. J. Wiffen, V. De Buffrénil, A. De Ricqlès, J.-M. Mazin, *Geobios* **28**, 625 (1995).
18. J. A. Massare, *J. Vertebr. Paleontol.* **7**, 121 (1987).
19. B. P. Kear, *Cretac. Res.* **26**, 769 (2005).
20. A. Longinelli, *Geochim. Cosmochim. Acta* **48**, 385 (1984).
21. Y. Kolodny, B. Luz, O. Navon, *Earth Planet. Sci. Lett.* **64**, 398 (1983).
22. Materials and methods are available as supporting material on Science Online.
23. T. F. Anderson, B. N. Popp, A. C. Williams, L. Ho, J. D. Hudson, *J. Geol. Soc. London* **151**, 125 (1994).
24. Y. Kolodny, B. Luz, in *Stable Isotope Geochemistry: A Tribute to Samuel Epstein*, H. P. Taylor, J. R. O'Neil, I. R. Kaplan, Eds. (Geochemical Society, University Park, St. Louis, MO, 1991), vol. 3, pp. 105–119.
25. Y. Kolodny, M. Raab, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **64**, 59 (1988).
26. C. Lécuyer *et al.*, *Paleoceanography* **18**, 1076 (2003).
27. Z. D. Sharp, V. Atudorei, H. Furrer, *Am. J. Sci.* **300**, 222 (2000).
28. A. Zazzo, C. Lécuyer, A. Mariotti, *Geochim. Cosmochim. Acta* **68**, 1 (2004).
29. M. R. Wadley, G. R. De Oliveira, E. J. Rohling, A. J. Payne, *Global Planet. Change* **32**, 89 (2002).
30. R. Amiot *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **243**, 412 (2007).
31. R. E. Barrick, A. G. Fischer, W. J. Showers, *Palaios* **14**, 186 (1999).
32. J. A. Massare, *Paleobiology* **14**, 187 (1988).
33. J. A. Massare, in *Mechanics and Physiology of Animal Swimming*, L. Maddock, Q. Bone, J. M. V. Rayner, Eds. (Cambridge Univ. Press, Cambridge, 1994), pp. 133–149.
34. P. Morrison, *Biol. Bull.* **123**, 154 (1962).
35. The authors thank M. Sander, B. McNab, and R. Seymour for preliminary discussions of these data during the workshop dedicated to "Sauropod gigantism" that was held in Bonn, Germany, in November 2008. We are also grateful to D. Brinkman and J. Gardner (Royal Tyrrell Museum), S. Etches and J. Clarke, G. Suan, J. Lindgren (Lund University), B. Kear (La Trobe University), and A. Schulp (Maastricht University) for providing samples. L. Simon and G. Escarguel helped us in the statistical treatment of data. This study was funded by both CNRS and the Institut Universitaire de France.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1379/DC1
Materials and Methods
Table S1

25 January 2010; accepted 28 April 2010
10.1126/science.1187443

Climate Change Will Affect the Asian Water Towers

Walter W. Immerzeel,^{1,2*} Ludovicus P. H. van Beek,² Marc F. P. Bierkens^{2,3}

More than 1.4 billion people depend on water from the Indus, Ganges, Brahmaputra, Yangtze, and Yellow rivers. Upstream snow and ice reserves of these basins, important in sustaining seasonal water availability, are likely to be affected substantially by climate change, but to what extent is yet unclear. Here, we show that meltwater is extremely important in the Indus basin and important for the Brahmaputra basin, but plays only a modest role for the Ganges, Yangtze, and Yellow rivers. A huge difference also exists between basins in the extent to which climate change is predicted to affect water availability and food security. The Brahmaputra and Indus basins are most susceptible to reductions of flow, threatening the food security of an estimated 60 million people.

Mountains are the water towers of the world (1), including for Asia, whose rivers all are fed from the Tibetan plateau and adjacent mountain ranges. Snow and glacial melt are important hydrologic processes in these areas (2, 3), and changes in temperature

and precipitation are expected to seriously affect the melt characteristics (4). Earlier studies have addressed the importance of glacial and snow melt and the potential effects of climate change on downstream hydrology, but these are mostly qualitative (4–6) or local in nature (7, 8). The relevance of snow and glacial melt for Asian river basin hydrology therefore remains largely unknown, as does how climate change could affect the downstream water supply and food security.

We examined the role of hydrological processes in the upstream areas, which we defined as

¹FutureWater, Costerweg 1G, 6702 AA Wageningen, Netherlands. ²Department of Physical Geography, Utrecht University, Post Office Box 80115, Utrecht, Netherlands. ³Deltares, Post Office Box 80015, 3508 TC Utrecht, Netherlands.

*To whom correspondence should be addressed. E-mail: w.immerzeel@futurewater.nl

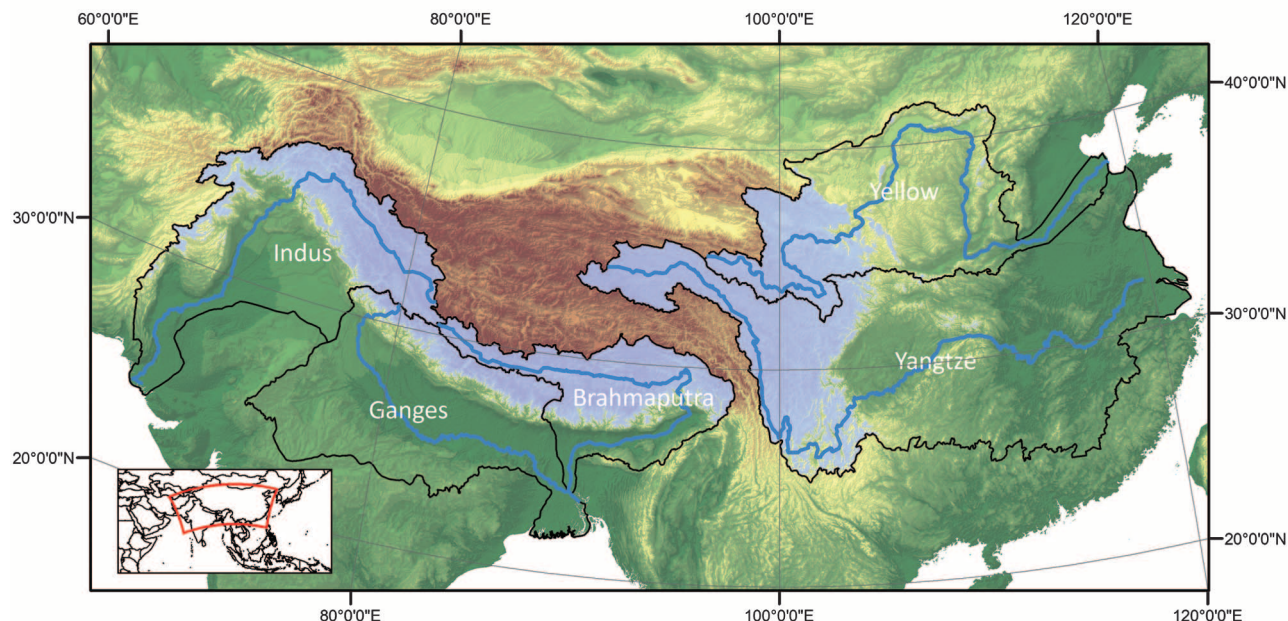


Fig. 1. Basin boundaries and river courses of the Indus, Ganges, Brahmaputra, Yangtze, and Yellow rivers. Blue areas denote areas with elevation exceeding 2000 masl. The digital elevation model in the background shows the topography ranging from low elevations (dark green) to high elevations (brown).

all areas higher than 2000 m above sea level (masl), on the water supply of the five major Southeast Asian basins (Fig. 1). These basins, which provide water to more than 1.4 billion people (over 20% of the global population), vary considerably in their characteristics (Table 1). The Yangtze has the largest population of the five basins, whereas the Ganges is the most densely populated. The Indus and Brahmaputra basins have extensive upstream areas (i.e., above 2000 m) and larger glaciated areas than the Yangtze and Yellow river basins (9). The Ganges, Brahmaputra and Yangtze basins are wetter than the Yellow and Indus basins (10). The Indus, Ganges, and Yangtze basins support large-scale irrigation systems (11) with high net irrigation water demand, but in the Indus the difference between basin precipitation and net irrigation demand is highest. We investigated three related components of these river basins: (i) the current importance of meltwater in overall river basin hydrology; (ii) observed cryospheric changes; and (iii) the effects of climate change on the water supply from the upstream basins and on food security.

We used the Normalized Melt Index (NMI) over the period 2001 to 2007 to quantify the importance of meltwater from the upstream areas on overall basin hydrology. NMI is defined as the volumetric snow and glacier upstream discharge divided by the downstream natural discharge. Upstream discharge is calculated with a calibrated snow melt runoff model (SRM) (12, 13). Downstream natural discharge is calculated by subtracting the natural evaporation (E_n) of the basins, calculated with a hydrological model (14), from precipitation (P) (15). E_n excludes additional evaporation from irrigated areas, because irrigation water is derived from upstream sources. The

Table 1. Characteristics of the five major Southeast Asian basins. Population data (2005) are based on the GPWv3 dataset [http://sedac.ciesin.columbia.edu/gpw (9 March 2009)]; precipitation data (average from 2001 to 2007) are based on (10); glacier areas are based on a dataset [http://glims.colorado.edu/glacierdata (9 March 2009)] provided by the Global Land Ice Measurements from Space (GLIMS) project (9). Irrigated areas and net irrigation water demand are based on (11). Upstream refers to the area > 2000 m.

Parameter	Indus	Ganges	Brahmaputra	Yangtze	Yellow
Total area (km ²)	1,005,786	990,316	525,797	2,055,529	1,014,721
Total population (10 ³)	209,619	477,937	62,421	586,006	152,718
Annual basin precipitation (mm)	423	1,035	1,071	1,002	413
Upstream area (%)	40	14	68	29	31
Glaciated area (%)	2.2	1.0	3.1	0.1	0.0
Annual upstream precipitation (%)	36	11	40	18	32
Annual downstream precipitation (%)	64	89	60	82	68
Irrigated area (km ²)	144,900	156,300	5,989	168,400	54,190
Net irrigation water demand (mm)	908	716	480	331	525

difference $P - E_n$ is therefore a measure of natural downstream discharge. The NMI is a more reliable measure than the commonly used meltwater fractions of total river discharge, which are affected by reservoirs, and water extractions (13). The great size of the basins that we analyze allows us to use melt parameters calculated for whole basins, rather than a different set of melt parameters for each different glacier, because each basin contains many glaciers of all types. Results from the NMI analysis (Fig. 2) indicate that for the present-day climate, meltwater plays an important role in the Indus and Brahmaputra river basins. This is most evident in the Indus: Discharge generated by snow and glacial melt is 151% of the total discharge naturally generated in the downstream areas. In the Brahmaputra basin this amounts to 27%. The contribution of snow and glacier water to the Ganges (10%), Yangtze (8%), and Yellow (8%) rivers is limited owing to comparatively large downstream

areas, limited upstream precipitation, smaller glaciers, and/or wet monsoon-dominated downstream climates (Table 1). In the Indus and Ganges basins, about 40% of the meltwater originates from glaciers, whereas in the other basins the glacial melt contribution is much less.

Since the end of the last ice age, an almost worldwide recession in glaciers has been observed (16), a trend that also applies to most of the glaciers in the Himalayas. Annual net imbalance rates of 0.5 to 0.9 m year⁻¹ have been observed from time series of digital elevation models in the Everest region in Nepal (17) and SPOT satellite imagery in the western Himalayas (18), whereas radioactivity analysis in ice cores revealed no net accumulation of ice in a high-elevation glacier in Tibet (19). However, there are some regional anomalies (13). We used the DMT-1 GRACE gravity model (20) in combination with derived precipitation trends (10) to identify large-scale

trends in snow and ice storage in each of the five basins. Results were inconclusive. We identified a negative trend of $-0.22 \pm 0.05 \text{ m year}^{-1}$ only in the Ganges basin. A positive trend of $0.19 \pm 0.02 \text{ m year}^{-1}$ was observed in the Indus basin, while in the other basins no discernable trends were identified (13). On the basis of this review, we conclude that there is a general decrease in the ice volumes of Asian basins, although regional anomalies exist and, as regional quantification of these trends is lacking, the uncertainty about these trends is substantial.

Fig. 2. Normalized melt index (NMI) for snow and glacier melt for the present (2000 to 2007) climate.

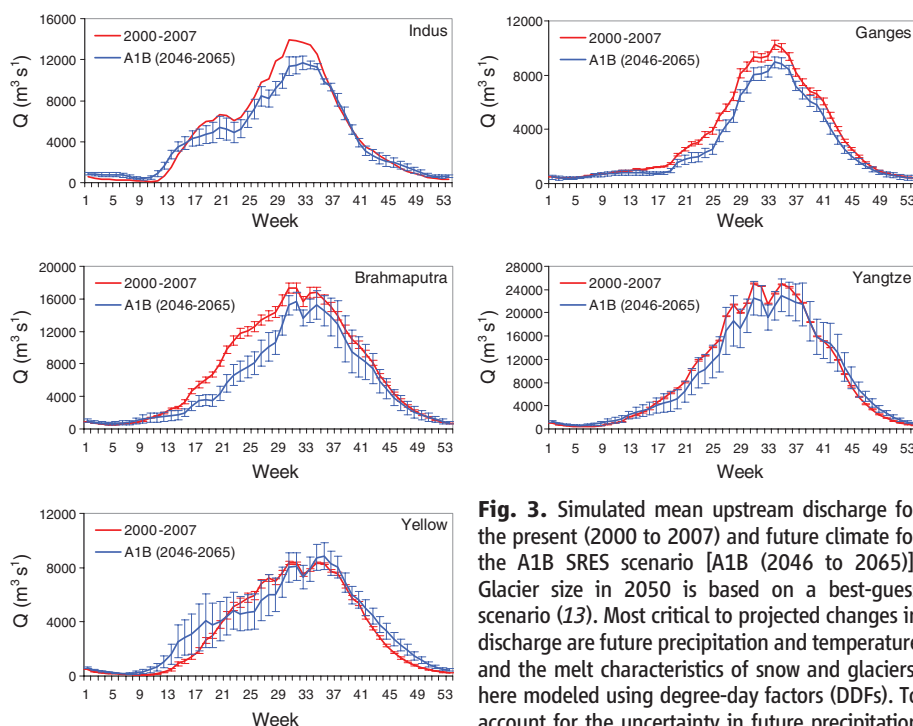
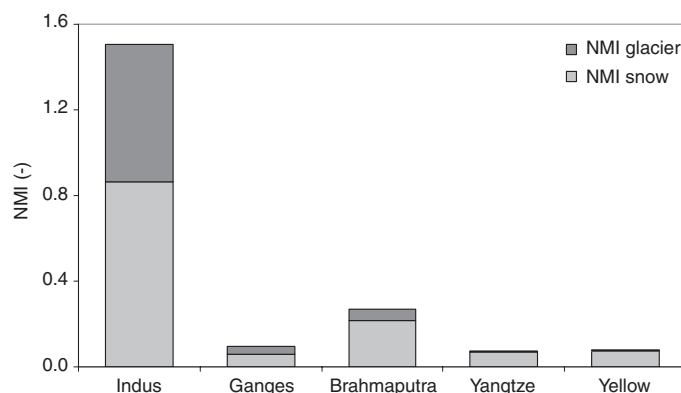


Fig. 3. Simulated mean upstream discharge for the present (2000 to 2007) and future climate for the A1B SRES scenario [A1B (2046 to 2065)]. Glacier size in 2050 is based on a best-guess scenario (13). Most critical to projected changes in discharge are future precipitation and temperature and the melt characteristics of snow and glaciers, here modeled using degree-day factors (DDFs). To account for the uncertainty in future precipitation and temperature, the best-guess scenario was run with five different GCMs (CCMA-CGCM3, GFDL-CM2, MPIM-ECHAM5, NIES-MIROC3, UKMO-HADGEM1). For the upper Indus basin, the ice and snow DDFs are constrained by observed runoff. No runoff observations were available for the other upper basins so that DDFs from the Indus were used. To account for this additional uncertainty, we performed a first-order-second-moment analysis (13). Vertical error bars ($\pm 1\sigma$) from the Indus thus include only the uncertainty about future climate, whereas the projected discharge of the other basins includes uncertainty about both future climate as well as basin-specific DDFs (assumed Gaussian with mean snow DDF = $4 \text{ mm } ^\circ\text{C}^{-1} \text{ day}^{-1}$, mean ice DDF = $7 \text{ mm } ^\circ\text{C}^{-1} \text{ day}^{-1}$, and both with $\sigma = 1 \text{ mm } ^\circ\text{C}^{-1} \text{ day}^{-1}$). Vertical error bars around present discharge are due to uncertainty about snow and ice DDFs only.

with five different GCMs (CCMA-CGCM3, GFDL-CM2, MPIM-ECHAM5, NIES-MIROC3, UKMO-HADGEM1). For the upper Indus basin, the ice and snow DDFs are constrained by observed runoff. No runoff observations were available for the other upper basins so that DDFs from the Indus were used. To account for this additional uncertainty, we performed a first-order-second-moment analysis (13). Vertical error bars ($\pm 1\sigma$) from the Indus thus include only the uncertainty about future climate, whereas the projected discharge of the other basins includes uncertainty about both future climate as well as basin-specific DDFs (assumed Gaussian with mean snow DDF = $4 \text{ mm } ^\circ\text{C}^{-1} \text{ day}^{-1}$, mean ice DDF = $7 \text{ mm } ^\circ\text{C}^{-1} \text{ day}^{-1}$, and both with $\sigma = 1 \text{ mm } ^\circ\text{C}^{-1} \text{ day}^{-1}$). Vertical error bars around present discharge are due to uncertainty about snow and ice DDFs only.

size were modeled (13): (i) a best guess based on glacier mass-balance calculations assuming trends in degree days and snowfall between current time and 2050 (calculated by the GCMs) to be linear, and (ii) an extreme (and unlikely) scenario with total disappearance of all glaciers to serve as a reference.

Upstream water supply is crucial to sustain upstream reservoir systems, which are used to store and release water to downstream areas when most needed. Irrigation water for the Indus Basin Irrigation Systems, which is the largest irrigation network in the world, is, for example, regulated through two major storage dams (Tarbela dam on the Indus River and the Mangla dam on the Jhelum River). Both are located in the upper Indus basin and are fed predominantly by meltwater. Any change in upstream water supply to these dams will have a profound effect on millions of people downstream. Our results show a substantial variation in changes in future water supply (Fig. 3). The best-guess glacier scenario resulted in a modeled decrease in mean upstream water supply from the upper Indus (-8.4%), the Ganges (-17.6%), Brahmaputra (-19.6%), and Yangtze rivers (-5.2%). Although these changes are considerable, they are less than the decrease in meltwater production would suggest, because this reduction is partly compensated for by increased mean upstream rainfall (Indus $+25\%$, Ganges $+8\%$, Brahmaputra $+25\%$, Yangtze $+5\%$, Yellow $+14\%$). The analysis even shows a notable 9.5% increase in upstream water yield in the Yellow River because this basin depends only marginally on glacial melt (Fig. 2). Results should be treated with caution, however, because most climate models have difficulty simulating mean monsoon and the interannual precipitation variation (21, 22), despite recent progress in improving the resolution of anticipated spatial and temporal changes in precipitation. Nevertheless, we conclude that although considerable cryospheric changes are to be expected, their impact will be less than anticipated by, for example, the 4AR of the Intergovernmental Panel on Climate Change (IPCC) (2). In that report, it was suggested that the current trends of glacier melt and potential climate change may cause the Ganges, Indus, Brahmaputra, and other rivers to become seasonal rivers in the near future. We argue that these rivers already are seasonal rivers, because the melt and rain seasons generally coincide and a decrease in meltwater is partially compensated for by an increase in precipitation. Figure 3 also shows a temporal shift in the upstream hydrograph patterns. The Yellow River, in particular, shows a consistent increase in early spring discharge. This is highly beneficial because most reservoirs are empty at the beginning of the growing season. An accelerated melt peak may thus alleviate a shortage of irrigation water in the drought-prone early stages of the growing season. In addition, further temporal shifts in the hydrograph could occur as a result of changes in seasonal glacier storage caused by a reduced glacier area (23).

Regardless of the compensating effects of increased rainfall in the two basins with the largest NMI, the Indus and the Brahmaputra, summer and late spring discharges are eventually expected to be reduced consistently and considerably around 2046 to 2065 after a period with increased flows due to accelerated glacial melt. Figure S2 also shows the extreme scenario in which all glaciers are assumed to have disappeared. Again, the Indus and Brahmaputra show the most pronounced changes.

These anticipated changes will also have considerable effects on food security. By relating changes in upstream water availability to net irrigation requirements (11), observed crop yields, caloric values of the crops, and required human energy consumption, one can estimate the change in the number of people that can be fed (13). The results (based on a best guess of 2050 glacier area) show a sizable difference between the five basins. Estimates range from a decrease of -34.5 ± 6.5 million people that can be fed in the Brahmaputra basin to -26.3 ± 3.0 million in the Indus basin, -7.1 ± 1.3 million in the Yangtze basin, and -2.4 ± 0.2 million in the Ganges basin, and an increase of 3.0 ± 0.6 million in the Yellow River basin. In total, we estimate that the food security of 4.5% of the total population will be threatened as a result of reduced water availability. The strong need for prioritizing adaptation options and further increasing water productivity is therefore ever more eminent (24). Figure S3 shows the projected changes in food security for the two different glacier scenarios. The difference between the scenarios is largest for the Indus and Brahmaputra basins, as a direct result of the high

NMI in these areas. Clearly, upstream discharge and downstream food security of the Indus and Brahmaputra basins are the most sensitive to climate change.

We conclude that Asia's water towers are threatened by climate change, but that the effects of climate change on water availability and food security in Asia differ substantially among basins and cannot be generalized. The effects in the Indus and Brahmaputra basins are likely to be severe owing to the large population and the high dependence on irrigated agriculture and meltwater. In the Yellow River, climate change may even yield a positive effect as the dependence on meltwater is low and a projected increased upstream precipitation, when retained in reservoirs, would enhance water availability for irrigated agriculture and food security.

References and Notes

1. D. Viviroli, H. H. Dürr, B. Messerli, M. Meybeck, R. Weingartner, *Water Resour. Res.* **43**, W07447 (2007).
2. R. V. Cruz *et al.*, in *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, M. L. Parry *et al.*, Eds. (Cambridge Univ. Press, Cambridge and New York, 2007).
3. W. W. Immerzeel, P. Droogers, S. M. de Jong, M. F. P. Bierkens, *Remote Sens. Environ.* **113**, 40 (2009).
4. T. P. Barnett, J. C. Adam, D. P. Lettenmaier, *Nature* **438**, 303 (2005).
5. B. C. Bates, Z. W. Kundzewicz, S. Wu, J. P. Palutikof, Eds., *Climate Change and Water* (IPCC, Geneva, 2008).
6. D. Cyranski, *Nature* **438**, 275 (2005).
7. P. Singh, L. Bengtsson, *J. Hydrol.* **300**, 140 (2005).
8. H. G. Rees, D. N. Collins, *Hydrol. Process.* **20**, 2157 (2006).
9. B. H. Raup *et al.*, *Global Planet. Change* **56**, 101 (2007).

10. G. J. Huffman *et al.*, *J. Hydrometeorol.* **8**, 38 (2007).
11. S. Siebert *et al.*, *Hydrol. Earth Syst. Sci.* **9**, 535 (2005).
12. J. Martinec, *Nord. Hydrol.* **6**, 145 (1975).
13. Materials and methods and discussion are available as supporting material on Science Online.
14. M. F. P. Bierkens, L. P. H. van Beek, *J. Hydrometeorol.* **10**, 953 (2009).
15. T. D. Mitchell, P. D. Jones, *Int. J. Climatol.* **25**, 693 (2005).
16. R. G. Barry, *Prog. Phys. Geogr.* **30**, 285 (2006).
17. T. Bolch, M. F. Buchroithner, T. Pieczonka, A. Kunert, *J. Glaciol.* **54**, 592 (2008).
18. E. Berthier *et al.*, *Remote Sens. Environ.* **108**, 327 (2007).
19. N. M. Kehrwalid *et al.*, *Geophys. Res. Lett.* **35**, L22503 (2008).
20. E. Klees *et al.*, *Geophys. J. Int.* **175**, 417 (2008).
21. H. Annamalai, K. Hamilton, K. R. Sperber, *J. Clim.* **20**, 1071 (2007).
22. S. Yang *et al.*, *J. Clim.* **21**, 3755 (2008).
23. P. Jansson, R. Hock, T. Schneider, *J. Hydrol. (Amst.)* **282**, 116 (2003).
24. D. B. Lobell *et al.*, *Science* **319**, 607 (2008).
25. This study was financially supported by the Netherlands Organisation for Scientific Research (NWO) through a CASIMIR grant (018 003 002) and by the European Commission (Call FP7-ENV-2007-1 grant 212921) as part of the CEOP-AEGIS project (www.ceop-aegis.org/) coordinated by the Université de Strasbourg. We thank P. Dittmar (Department of Earth Observation and Space Systems, Technical University of Delft) for assistance with the GRACE trend analysis.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1382/DC1

Materials and Methods

SOM Text

Figs. S1 to S4

Table S1

References

12 October 2009; accepted 6 May 2010

10.1126/science.1183188

Periodic, Chaotic, and Doubled Earthquake Recurrence Intervals on the Deep San Andreas Fault

David R. Shelly

Earthquake recurrence histories may provide clues to the timing of future events, but long intervals between large events obscure full recurrence variability. In contrast, small earthquakes occur frequently, and recurrence intervals are quantifiable on a much shorter time scale. In this work, I examine an 8.5-year sequence of more than 900 recurring low-frequency earthquake bursts composing tremor beneath the San Andreas fault near Parkfield, California. These events exhibit tightly clustered recurrence intervals that, at times, oscillate between ~ 3 and ~ 6 days, but the patterns sometimes change abruptly. Although the environments of large and low-frequency earthquakes are different, these observations suggest that similar complexity might underlie sequences of large earthquakes.

The idea that fault segments rupture in similar quasi-periodic earthquakes is a model increasingly incorporated into earthquake forecasts (1, 2). In certain places,

however, large systematic deviations from average recurrence intervals (3, 4) call into question the utility of this simple model. Examples of complex earthquake recurrence include chaotic intervals as well as period doubling, which is usually transitional between periodic and chaotic behavior (5, 6).

Numerical models sometimes show complexity in earthquake recurrence intervals, such as chaotic behavior and period doubling, in addition to periodic slip (7–13). Laboratory studies have also demonstrated evidence for complexity in frictional sliding experiments, especially near the transition from stable sliding (fault creep) to stick-slip (earthquakes) (5, 9, 14). Large natural earthquakes may be subject to similar controls, with fault interactions accounting for variations in recurrence intervals on a given fault segment (15, 16), but tens to hundreds of years between large events on a given natural fault strongly limit the number of observable intervals. Even in rare instances where long records exist, such as Nankai Trough, Japan (16), a 1300-year record of eight events on a given segment probably does not reveal the full variability of earthquake rupture. Likewise, the timing of the 2004 moment magnitude (M_w) 6.0 Parkfield, California earthquake was not anticipated based on this segment's history of six prior earthquakes dating back to 1857 (17).

Smaller earthquakes, with correspondingly shorter recurrence times, provide a more tractable natural laboratory for investigating recurrence behavior. In this study, I examined the recurrence intervals of a “family” of similar low-frequency

U.S. Geological Survey, Menlo Park, CA 94025, USA. E-mail: dshelly@usgs.gov

earthquakes (LFEs), which occur as part of a tectonic tremor beneath the San Andreas fault (SAF) in central California near Parkfield. Tremors on the SAF occur below the seismogenic zone near the base of the crust (18, 19), probably by shear slip (20), and are composed of a sequence of tiny overlapping individual events (i.e., LFEs) (21).

The LFE family analyzed here occurs in short bursts, as typically two to six similar (and probably adjacent) events in rapid succession over a span of 1 to 2 min, followed by 2 to 7 days of quiescence. Recurrence intervals are times between these bursts; events within a given burst are those with interevent times less than 0.01 days (14.4 min). LFEs were identified and classified as a family via cross-correlation with a waveform template (22). Relative ground velocity amplitudes of these events are also measured (23).

This particular family is temporally isolated in its occurrence—that is, it is not preceded or followed by a detectable tremor in other families—suggesting that events occur on a spatially isolated patch on the fault. Other LFE families show many elements of the behavior documented here but appear to interact more with each other (22). The best-fitting location (Fig. 1), based on stacked waveforms at 31 stations (fig. S1), is a few kilometers deeper than most other LFE families (22), with an estimated depth of 29.75 km (near the crust-mantle boundary). The best-fitting epicentral location offsets ~6 km to the northeast from the fault surface trace. This suggests that these events occur on a dipping or offset fault; however, given the orientation of location uncertainties (Fig. 1) and unknown seismic velocity structure at this depth, a location directly beneath the surface trace is still possible.

From mid-2001 (when observations began) through 2002, recurrence intervals vary between 2 and 7 days, with most intervals between 2 and 3.5 days (Fig. 2). Intervals greater than 7 days (and perhaps those of 5 to 7 days) during this time period may be indicative of missed events, because station noise levels were higher before station changes in mid-to-late 2003. In early to mid-2003, before station modifications, the recurrence pattern abruptly changes: The scatter around 3 days decreases, tightening near 2.8 days, and a second period of ~5.9 days emerges. While oscillating between these intervals over the next several months, the ~6-day period gains in proportion (fig. S3). About the time of the 2003 M_w 6.5 San Simeon earthquake, 80 km to the west, the balance between ~3- and ~6-day intervals becomes more even, and the ~3-day period even dominates for a short time in mid-2004. The 3- to 6-day oscillation resumes in late 2004 up until the M_w 6.0 Parkfield earthquake, bearing similarity to period doubling reported in laboratory experiments (5).

Despite a separation of 30 km from the coseismic rupture zone of the Parkfield earthquake, recurrence behavior changes dramatically after the earthquake. The shorter ~3-day interval abruptly drops to ~2.5 days, and it drops further to ~2 days over the next 2 months. This could result from the arrival of a creep front from the earthquake, or

a change in partitioning between seismic and aseismic slip (24, 25). The very small amplitudes during this time also suggest weakening and repeated failure of this fault patch with lower accumulated stress, perhaps induced by strong shaking from the earthquake. Over the next 5 years, this dominant interval slowly recovers back toward 3.0 days.

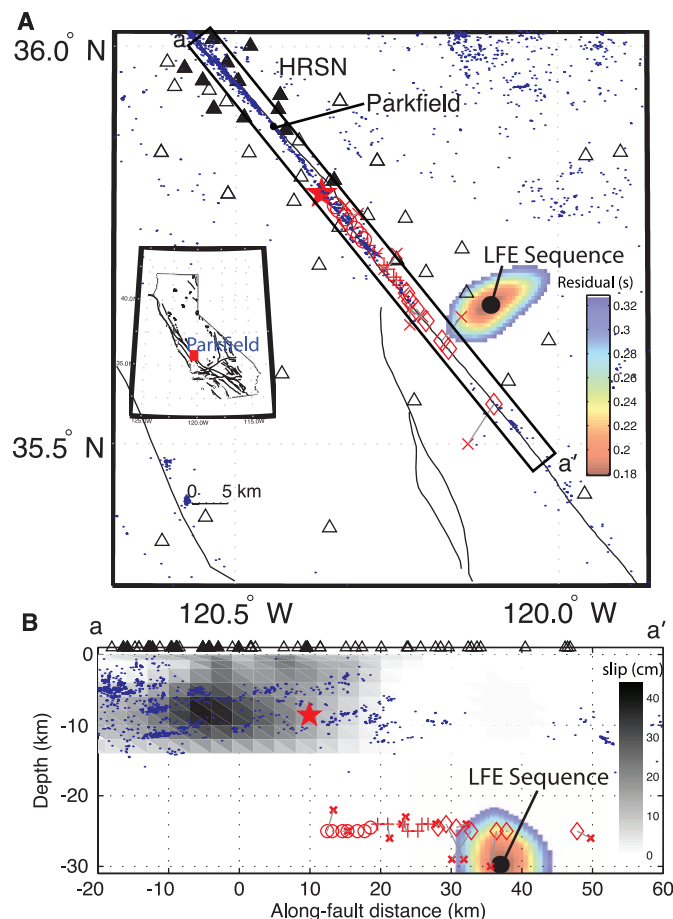
In addition to the overall decrease in recurrence interval, the well-organized longer period disappears after the earthquake. A few intervals of 4 to 5.5 days may be shortened versions of the 6-day interval (or they might result from events missed during the aftershock sequence), but they lack the prevalence and the well-defined interval from before the earthquake. The single 7-day interval immediately after the earthquake may be due to one or more bursts lost in the active aftershock sequence. Intervals approaching 6 days return in mid-2006, yet they are less prevalent and lack the tight interval seen from 2003 to 2004. From late 2008 through mid-2009, the more abundant intervals of 5 to 6 days suggest a transition back to dual-period behavior. In mid-2009, however, the pattern changes again: The 6-day events disappear, and chaotic variability, with intervals ranging from 2 to 5 days, possibly emerges.

Multiple lines of evidence suggest that the 6-day intervals observed for this LFE family are not

simply an artifact of a missed event. The abrupt loss of the 6-day period after the Parkfield earthquake occurred during a time when missed events should, if anything, be more common due to the concurrent aftershock sequence. Additionally, if missed events are common, we should expect to also see apparent 9-day intervals (corresponding to two missed events), but such events are absent from late 2003 onward. There is also the tendency for larger amplitudes after 6-day intervals (Fig. 2 and fig. S2), consistent with the expected greater accumulated stress over a longer time period. Furthermore, 6-day events have fewer events per burst (fig. S2), suggesting that additional stress allows this fault patch to fail in fewer, larger LFEs compared with 3-day events. In contrast, the amplitude and number of events per burst show little correlation with the subsequent interval (fig. S2). This is suggestive of a model of stress dropping to a nearly constant value after each burst, rather than a fixed failure stress. In this sense, these events are more slip-predictable than time-predictable (26).

Assuming that events cumulatively approximate the fault slip rate of ~3 cm/year and slip is roughly proportional to the time since the previous event, 3- and 6-day events represent ~0.25 and 0.5 mm of slip, respectively. Because some slip could occur aseismically between events (24, 25), these slip values represent upper bounds.

Fig. 1. Map (A) and cross section (B) showing best-fit location of the LFE sequence examined in this work (filled black circle). The color scale shows corresponding mean residuals at nearby locations (map view is projected at a best-fit depth of 29.75 km). Red x symbols indicate best-fitting single LFE locations located from *P*- and *S*-wave arrival times; circles, crosses, and diamonds represent inferred locations (28); the red star denotes the hypocenter of the 2004 M_w 6.0 Parkfield earthquake; and the shaded region in (B) shows associated coseismic and postseismic (first 230 days) slip (29). The rectangular box and labels *a* and *a'* in (A) indicate the limits and orientation of the cross section in (B). Solid black triangles indicate High Resolution Seismic Network (HRSN) borehole seismic stations used for detection, open triangles denote other seismic stations used for location (fig. S1), and blue dots show relocated earthquakes from 1984 to 2003 (30). Modified from (28).



Considering the subsequent recurrence interval versus the previous interval for each event provides further insight into this behavior (Fig.

3). Along the diagonal are locally periodic events with identical previous and subsequent intervals. Since 2001, locally periodic behavior occurred at

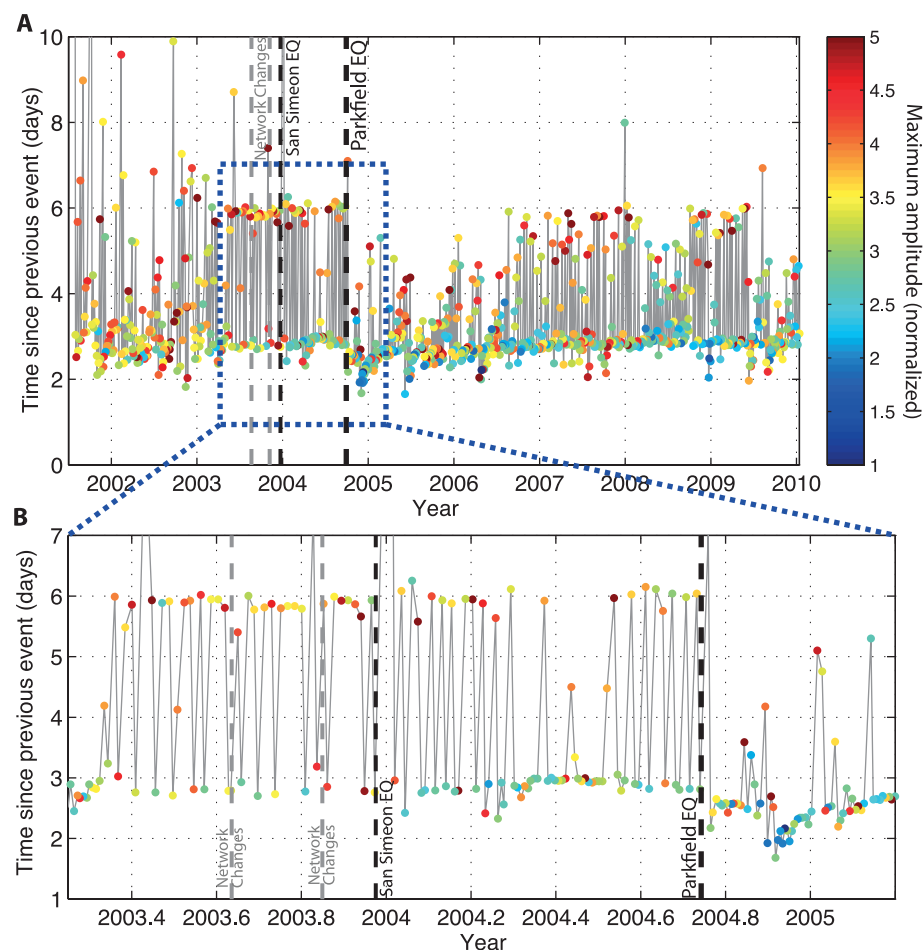
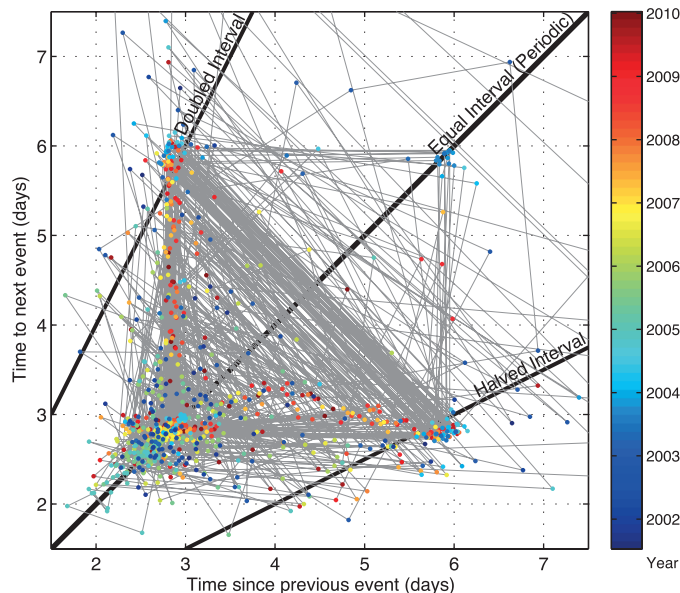


Fig. 2. Recurrence interval versus time for LFE family bursts. **(A)** Recurrence history, mid-2001 to 2010. The color scale indicates the maximum amplitude (ground velocity) of each burst, normalized to the first-percentile amplitude of all events in this family. Solid gray lines connect consecutive events. **(B)** Zoomed-in view of section in (A), highlighting the changing behavior after the 2004 M_w 6 Parkfield earthquake.

Fig. 3. Subsequent interval versus preceding interval for each event. Points are color-coded with time, and gray lines connect consecutive points. Events with equal subsequent and preceding intervals are plotted on the diagonal. The clear off-diagonal structure indicates common oscillations between periods of ~3 and ~6 days.



intervals of 2.4 to 3.0 days and (dominantly in late 2003) 5.8 to 6.0 days. On the other hand, points off the diagonal indicate deviations from periodic behavior. Particularly prevalent are paths between 3- and 6-day events, indicating oscillation in intervals between 3 and 6 days. Asymmetry between preceding and subsequent events in this plot reflects directionality of the system in time. For instance, if perturbed from periodic occurrence, the sequence sometimes takes several cycles to slowly return to the dominant period, which behaves much like a stable attractor (5).

As suggested by laboratory and numerical studies, multiple mechanisms could explain the oscillating recurrence period. One possibility is that the frictional properties themselves regulate this behavior. This concept may have relevance here, as the tremor activity suggests that this region is transitional between stick-slip and stable sliding, precisely the regime where period doubling has been observed in the laboratory (5). Alternatively, interaction between neighboring fault patches could produce similar patterns (11). For example, an adjacent aseismic patch could slip periodically every 3 days, often triggering the LFE patch. If the aseismic patch does not trigger an LFE, then 3 more days might pass before the LFE is triggered.

The LFEs examined here occur deeper on the fault than large earthquakes, and different frictional properties may govern them. In particular, surrounding temperatures and fluid pressures are probably higher for the LFEs. Notably, these events may have more in common with large, slow-slip events, which are nearly periodic in some locations (27). Nevertheless, the LFEs' frequent occurrence provides an opportunity to examine fault slip recurrence over hundreds of cycles, which was previously possible only in laboratory or numerical simulations. The recurrence patterns of the highlighted event family suggest that complexity previously documented in laboratory and numerical simulations may play an important role in the real earth. For example, large, nonrandom deviations from periodic behavior may interrupt sequences of nearly periodic events (fig. S4). In particular, it appears that small changes in conditions can produce dramatic changes in behavior, a trait characteristic of complex systems. Although the observed history of earthquakes may remain the best available guide to a fault's future events, the LFE sequence analyzed here suggests possible limitations in extrapolating a short record of earthquake recurrence.

References and Notes

1. E. H. Field et al., *Bull. Seismol. Soc. Am.* **99**, 2053 (2009).
2. K. M. Scharer et al., *Geology* **38**, 555 (2010).
3. W. L. Ellsworth et al., "A physically-based earthquake recurrence model for estimation of long-term earthquake probabilities," U.S. Geological Survey Open-File Report 99-522 (U.S. Geological Survey, Reston, VA, 1999).
4. R. Weldon, K. Scharer, T. Fumal, G. Biasi, *GSA Today* **14**, 4 (2004).
5. Y. Gu, T.-F. Wong, in *Nonlinear Dynamics and Predictability of Geophysical Phenomena*, *Geophysical Monograph*, vol. 83, W. I. Newman et al., Eds. (American Geophysical Union, Washington, DC, 1994), pp. 15–35.

6. B. R. Julian, *J. Volcanol. Geotherm. Res.* **101**, 19 (2000).
7. R. Burridge, L. Knopoff, *Bull. Seismol. Soc. Am.* **57**, 341 (1967).
8. A. L. Ruina, *J. Geophys. Res.* **88**, 10359 (1983).
9. J. Huang, D. L. Turcotte, *Geophys. Res. Lett.* **17**, 223 (1990).
10. J. R. Rice, *J. Geophys. Res.* **98**, 9885 (1993).
11. S. Ma, C. He, *Tectonophysics* **337**, 135 (2001).
12. T. W. Becker, *Geophys. Monogr.* **120**, 5 (2001).
13. B. Erickson, B. Birnir, D. Lavallée, *Nonlinear Processes Geophys.* **15**, 1 (2008).
14. J. D. Weeks, T. E. Tullis, *J. Geophys. Res.* **90**, 7821 (1985).
15. K. Ito, *J. Geophys. Res.* **85**, 1399 (1980).
16. J. Huang, D. L. Turcotte, *Nature* **348**, 234 (1990).
17. W. H. Bakun *et al.*, *Nature* **437**, 969 (2005).
18. R. M. Nadeau, D. Dolenc, *Science* **307**, 389 (2005); published online 9 December 2004 (10.1126/science.1107142).
19. D. R. Shelly *et al.*, *Geophys. Res. Lett.* **36**, L01303 (2009).
20. A. Thomas, R. M. Nadeau, R. Bürgmann, *Nature* **462**, 1048 (2009).
21. D. R. Shelly, G. C. Beroza, S. Ide, *Nature* **446**, 305 (2007).
22. D. R. Shelly, *Nature* **463**, 648 (2010).
23. Materials and methods are available as supporting material on Science Online.
24. N. M. Beeler, D. L. Lockner, S. H. Hickman, *Bull. Seismol. Soc. Am.* **91**, 1797 (2001).
25. T. Chen, N. Lapusta, *J. Geophys. Res.* **114**, B01311 (2009).
26. K. Shimazaki, T. Nakata, *Geophys. Res. Lett.* **7**, 279 (1980).
27. G. Rogers, H. Dragert, *Science* **300**, 1942 (2003); published online 8 May 2003 (10.1126/science.1084783).
28. D. R. Shelly, *Geophys. Res. Lett.* **36**, L17318 (2009).
29. J. Murray, J. Langbein, *Bull. Seismol. Soc. Am.* **96**, S283 (2006).
30. F. Waldhauser, D. P. Schaff, *J. Geophys. Res.* **113**, B08311 (2008).
31. The HRSN borehole network is operated by the University of California, Berkeley. Data were obtained through the Northern California and Southern California Earthquake Data Centers. I thank A. Michael, N. Beeler, R. Bürgmann, G. Beroza, P. Johnson, E. Daub, and R. Guyer for their helpful feedback and discussion.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1385/DC1

Materials and Methods

Figs. S1 to S4

References

17 March 2010; accepted 3 May 2010

10.1126/science.1189741

Stochastic Community Assembly Causes Higher Biodiversity in More Productive Environments

Jonathan M. Chase

Net primary productivity is a principal driver of biodiversity; large-scale regions with higher productivity generally have more species. This pattern emerges because β -diversity (compositional variation across local sites) increases with productivity, but the mechanisms underlying this phenomenon are unknown. Using data from a long-term experiment in replicate ponds, I show that higher β -diversity at higher productivity resulted from a stronger role for stochastic relative to deterministic assembly processes with increasing productivity. This shift in the relative importance of stochasticity was most consistent with the hypothesis of more intense priority effects leading to multiple stable equilibria at higher productivity. Thus, shifts in community assembly mechanisms across a productivity gradient may underlie one of the most prominent biodiversity gradients on the planet.

Variation in the net primary productivity of a given ecosystem (e.g., the rate of carbon fixation through photosynthesis per unit area) plays a fundamental role in driving variation in biodiversity of both plants and animals across regions; more productive regions typically have higher levels of biodiversity (1–4). This positive productivity-biodiversity relationship typically manifests when biodiversity is measured at relatively large spatial scales; when biodiversity is measured at smaller spatial scales, the pattern is weaker and more variable (4, 5). The discrepancy between productivity's effect on biodiversity at smaller and larger spatial scales can be resolved by recognizing that site-to-site variability in species composition, known as β -diversity, often increases with increasing productivity (6–12). However, the mechanisms leading to the increase in β -diversity with increasing productivity remain largely unknown; most studies to date have been correlational. The factors that cause variation in β -diversity are among the most important, but poorly understood, influences on global variation in biodiversity (13–15).

β -diversity can arise from community assembly mechanisms involving (i) purely deterministic processes when habitat heterogeneity creates different niches across localities to which different groups of species are favored; (ii) purely stochastic processes including ecological drift, dispersal limitation, and differential colonization/extinction dynamics across localities; or (iii) the interaction between stochastic and deterministic processes when stochastic variation in the history of colonization leads to more deterministic priority effects that vary across localities (16). To account for the generally observed pattern of increasing β -diversity with increasing productivity (6–12), one or more of these mechanisms must increase in its relative importance with increasing productivity. Deterministic processes could lead to increased β -diversity with productivity if the magnitude of variation in productivity among localities generally increases as the mean productivity of a region increases. However, studies that have specifically examined correlations between the magnitude of among-site heterogeneity and average productivity in a given region, and how that might influence β -diversity, have found no such relationships (6, 8, 17). As a result, it is often more likely that variation in the importance of stochastic processes, or in the interaction between stochastic

and deterministic processes, leads to the observed increase of β -diversity with productivity.

There are two distinct mechanisms by which the relative importance of stochasticity in community assembly, leading to high β -diversity, can increase with productivity. First, stochastic variation in colonization or extinction among localities can result in ecological drift and dispersal limitation (18), or temporally variable metacommunities with locally unstable dynamics (19), both of which can create high β -diversity among localities. This spatiotemporal variability in community structure (β -diversity) may increase in frequency with increasing productivity if rates of colonizations or extinctions increase, or if interspecific interactions leading to unstable dynamics increase, with productivity (20). Second, stochasticity in colonization history can lead to priority effects, which can then deterministically create multiple stable equilibria of community structure in different localities, leading to high β -diversity [e.g., (16)]. There are at least two related theoretical reasons as to why priority effects leading to multiple stable equilibria might be more frequent in higher-productivity regions: (i) If a smaller proportion of the species in the species pool can persist in lower productivity, and if that smaller proportion is relatively nested within the group of species that can persist in higher productivity, at least in the absence of interspecific interactions, priority effects leading to multiple stable equilibria can be more frequent at higher productivity from simple probability (21). (ii) In a food web where multiple species share common resources and common predators, increasing the level of productivity can create a saddle between dominance by species that are more effective competitors for shared resources and dominance by species that are more effective apparent competitors mediated through shared predators. This can lead to a higher propensity for multiple stable equilibria at high productivity, depending on which species with which traits (resource or apparent competition specialists) colonize a given community first (22). Similar effects should emerge under different combinations of traits and species interactions, as long as the balance of a species' interspecific effects on other species in the community becomes more likely to exceed its intra-

Department of Biology and Tyson Research Center, Washington University, St. Louis, MO 63130, USA. E-mail: jchase@wustl.edu

specific effects as the level of primary productivity increases.

Here, I present the results from a 7-year experiment in 45 experimental freshwater ponds arrayed in an old-field at Washington University's Tyson Research Center, near Saint Louis, Missouri (21). This experiment was used to explicitly address (i) the relative contribution of stochastic processes (mediated through differential colonization history, ecological drift, or colonization/extinction dynamics) to within-treatment β -diversity, and the resulting effect of productivity on species diversity at local and regional spatial scales; and (ii) the potential mechanisms by which variation in the importance of these stochastic processes across productivity treatments manifests—through asynchronous temporal variation, or through priority effects leading to possible multiple stable equilibria. I treated ponds with three levels of nutrients to stimulate productivity and replicated each treatment with 15 ponds to observe any differences in within-treatment β -diversity. To create stochasticity in colonization, allowing for possible priority effects, I randomly varied the order of introduc-

tion of several species of producers and animals among each replicate pond over the first 2 years of the experiment; many other species colonized on their own accord, albeit somewhat stochastically (see table S1 for a list of species and whether they were introduced or colonized on their own). After these initial colonizations, stochastic processes (colonization and extinction dynamics, ecological drift, dispersal limitation) and deterministic processes (species sorting, priority effects) occurred naturally to influence patterns of β -diversity over the final 5 years of the experiment.

At the local (within-pond) scale, there were no significant effects of the productivity treatments on the richness of producers or animals within a given pond (α -diversity) [Producer analysis of variance (ANOVA): $F_{2,44} = 1.21$, $P > 0.31$; Animal ANOVA: $F_{2,44} = 1.48$, $P > 0.25$]. However, using pairwise dissimilarity metrics [1-Jaccard's (incidence-based), 1-Bray-Curtis (abundance-based)] as a proxy for β -diversity, I found considerably and consistently high β -diversity among replicate high-productivity ponds, intermediate levels of β -diversity among replicate intermediate-productivity

ponds, and low β -diversity among replicate low-productivity ponds (Fig. 1 and table S2). Although similar results have been seen in correlational studies (6–12), the few experimental studies that have been examined in this context (17, 23) have not been able to quantify whether higher β -diversity among replicates in higher-productivity habitats was a stable phenomenon, nor could they discern the mechanisms underlying this effect.

I found that only a subset of the total species pool could persist in the lowest-productivity treatment, imprinting a strong environmental filter on community assembly. Alternatively, the majority of species in the species pool could persist, at least on occasion, in the highest-productivity treatment, suggesting a much smaller role for environmental filtering in this treatment; ponds in the intermediate-productivity treatment were intermediate in the strength of the environmental filter. Specifically, 21 species of macroscopic producers (vascular plants and filamentous algae and cyanobacteria) were recorded from at least one experimental pond (no attempts were made to identify unicellular algae) (table S1). Of those species, 90% (19 of 21) were observed at least once in the highest-productivity treatments, whereas only 33% (7 of 21) were observed in the lowest-productivity treatments. Although too numerous to list, producer species that were able to persist in the lowest-productivity ponds included the vascular macrophyte *Elodea canadensis* and the macrophyte-like algae *Chara vulgaris*, both of which are known to be common in nutrient-poor waters. Likewise, of the 77 species of animals observed in at least one experimental pond (table S1), 88% (68 of 77) were observed at least once in the highest-productivity ponds, whereas only 39% (30 of 77) were observed at least once in the lowest-productivity ponds. Species able to persist in low-productivity ponds included small snails (e.g., *Gyraulus parvus*, *Physella gyrina*) and mayflies (*Callibaetis* spp.), which are species that appear to experience a trade-off between resource acquisition and risk of predation from larger predatory species (e.g., dragonfly larvae, hemipteran water bugs) (24) that were more likely to occur in the higher-productivity treatments.

Because the group of species that persisted in at least one pond of the low-productivity treatment was almost entirely a nested subset of the group of species that persisted in at least one pond of the high-productivity treatment, there was scale dependence in the relationship between productivity and species richness [as has been observed in surveys of natural ponds (6)]. That is, although neither producer nor animal richness was influenced by productivity treatment at the scale of the local pond (see above), at the scale of the entire treatment (γ -diversity) ($N = 15$ replicates), producer richness increased by nearly a factor of 3 (7 to 19 species) and animal richness increased by more than a factor of 2 (30 to 68 species) from low to high productivity.

If higher-productivity ponds had more individuals than lower-productivity ponds, one possible

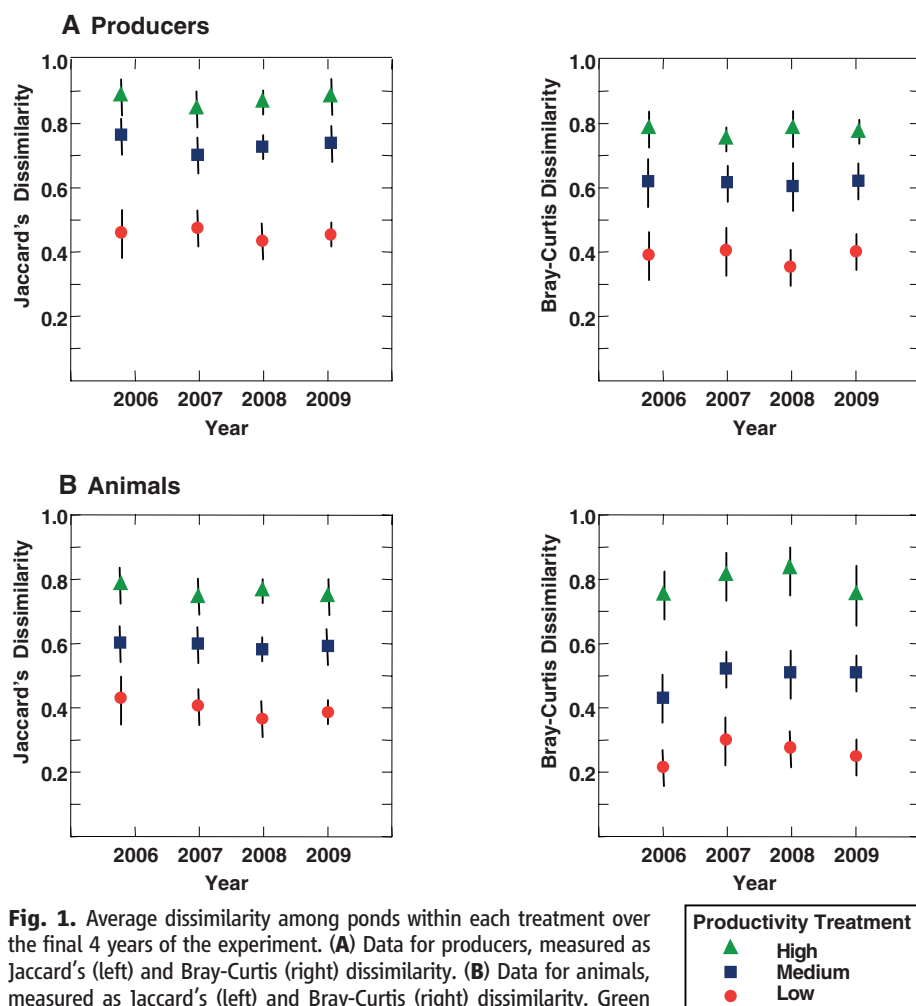


Fig. 1. Average dissimilarity among ponds within each treatment over the final 4 years of the experiment. (A) Data for producers, measured as Jaccard's (left) and Bray-Curtis (right) dissimilarity. (B) Data for animals, measured as Jaccard's (left) and Bray-Curtis (right) dissimilarity. Green triangles, blue squares, and red circles indicate the average dissimilarity among replicates within high-, intermediate-, and low-productivity treatment, respectively. Error bars represent 1 SE. Because these data are nonindependent, significance tests for treatment and time effects require a randomization procedure (21), results of which are presented in table S2.

explanation for the increased species richness observed in the high-productivity ponds could be that rare species were simply easier to detect because of a sampling effect. However, this was not likely to be the case for several reasons. First, as part of my sampling methods, I specifically targeted rare species to ensure adequate sampling (21). Second, rarefying local species richness did not change the lack of a significant relationship between productivity and local species richness (Producer ANOVA, $F_{2,44} = 2.27$, $P > 0.12$; Animal ANOVA, $F_{2,44} = 2.30$, $P > 0.13$). Finally, because the relative abundance of species was highly variable among replicates within a treatment, rarefying richness across the entire treatment to account for differences in abundance had no influence on the richness observed at the scale of the entire treatment, nor on β -diversity among replicates within a treatment. That said, because there was no effect of productivity on local richness but a large effect of productivity on the number of species observed in the entire treatment, the species in the high-productivity treatments had lower

occupancy rates on average, and thus were regionally rarer, than those in the lower-productivity treatments. Because I observed consistently higher β -diversity among otherwise identically treated ponds in the high-productivity treatments relative to the low-productivity treatments, this suggests that stochastic processes were likely to play a stronger role in the assembly of higher-productivity communities. However, most measures of β -diversity are influenced by changes in the ratios of the numbers of species that live locally and regionally (e.g., $\alpha:\gamma$). Thus, to tease apart the relative importance of stochastic versus deterministic mechanisms underlying community assembly along this experimental productivity gradient, I performed a null model analysis (21) on the community-level data from the final year of the experiment. I found that the observed β -diversity (measured as Jaccard's dissimilarity) among high-productivity communities was indistinguishable from the null expectation that included only stochastic community assembly processes [permutational

analysis of multivariate dispersions (PERMDISP) comparing the observed dissimilarity relative to the average dissimilarity produced by the null model: Producers, $F_{1,28} = 1.35$, $P > 0.4$; Animals, $F_{1,28} = 0.59$, $P > 0.8$]. Alternatively, the degree of dissimilarity observed in both intermediate-productivity communities (PERMDISP: Producers, $F_{1,28} = 3.18$, $P < 0.04$; Animals, $F_{1,28} = 5.59$, $P < 0.03$) and low-productivity communities (PERMDISP: Producers, $F_{1,28} = 11.35$, $P < 0.01$; Animals, $F_{1,28} = 15.54$, $P < 0.01$) was significantly lower than the null expectation, indicating that some degree of determinism likely structured these communities. To more directly test the deviations from the null expectation among productivity treatments, I calculated a probability distribution of deviations from the null-expected number of shared species among pairwise community comparisons (21). I found that the values of this metric were larger, and thus less deviant from the null expectation, among replicates of higher-productivity ponds relative to replicates of lower-productivity ponds; replicates of intermediate-productivity ponds were intermediate in their deviation from the null expectation (PERMDISP: Producers, $F_{2,42} = 43.50$, $P < 0.001$; Animals, $F_{2,42} = 132.59$, $P < 0.001$; all pairwise comparisons $P < 0.02$) (Fig. 2). This confirms an increasing role for stochastic assembly processes leading to higher β -diversity with increasing productivity.

Above, I showed that the magnitude of β -diversity within treatments was consistent over the last 4 years of the experiment. However, this does not indicate whether the observed variation in β -diversity emerged because of higher asynchronous temporal variability in community structure (18–20) or because of more intense priority effects potentially leading to more frequent multiple stable equilibria (16). By measuring the rates of change in community composition within each experimental pond over the final 4 years of the experiment (21), I was able to test whether asynchronous temporal variability could be implicated or eliminated as a possible underlying mechanism behind the observed patterns of β -diversity. Perhaps surprisingly, given the large amount of spatial variability in community composition (i.e., β -diversity) seen across the array of experimental ponds, and particularly in the high-productivity treatment, there was little temporal change in the structure of the communities within a given pond from year to year, and there were no systematic differences in the rates of temporal change within ponds among the productivity treatments (Table 1). These results are inconsistent with hypotheses of purely stochastic community assembly (e.g., ecological drift and dispersal limitation) or unstable local community dynamics (18–20), and are more consistent with the hypothesis that stochastic colonizations in higher-productivity ponds lead to more frequent priority effects and multiple stable equilibria (16). The relative temporal uniformity of community structure within the ponds across 4 years encompassed tens to hundreds of generations of a majority of organisms in these

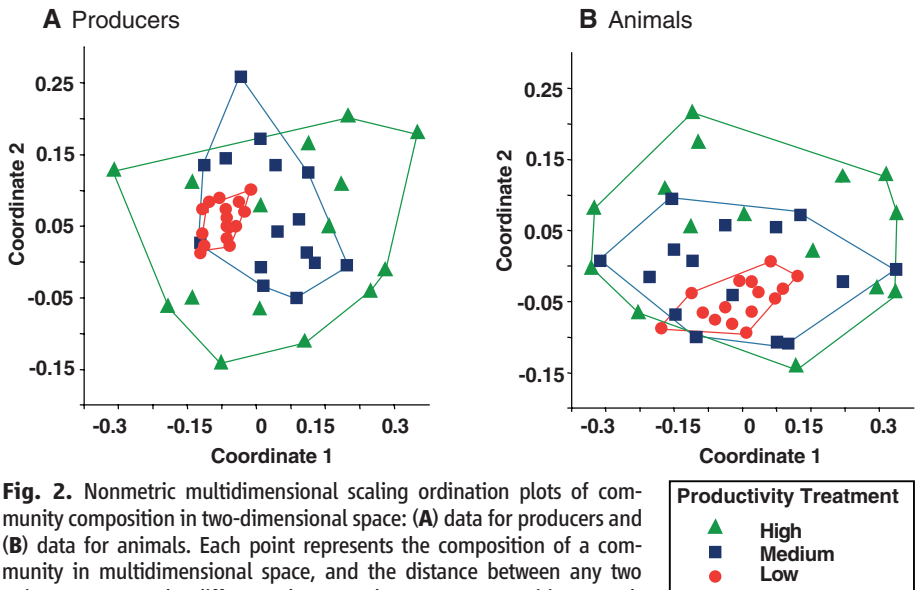


Fig. 2. Nonmetric multidimensional scaling ordination plots of community composition in two-dimensional space: (A) data for producers and (B) data for animals. Each point represents the composition of a community in multidimensional space, and the distance between any two points represents the difference between those two communities according to a modified Raup-Crick dissimilarity metric, which indicates the degree to which the null-expected number of shared species between any two communities deviates from the observed number of shared species (21). Communities that are closer together are more deviant from the null expectation, whereas communities that are farther apart are less deviant from the null expectation. Symbol shapes and colors are as in Fig. 1. Lines represent the minimum convex hulls around the data.

Table 1. Comparison of the average within-pond dissimilarity among years (2006 to 2009) in each treatment.

Productivity treatment	Jaccard's dissimilarity		Bray-Curtis dissimilarity	
	Producer	Animal	Producer	Animal
Low	0.03 ± 0.03	0.19 ± 0.13	0.18 ± 0.05	0.24 ± 0.13
Medium	0.02 ± 0.02	0.13 ± 0.12	0.15 ± 0.07	0.27 ± 0.17
High	0.02 ± 0.03	0.16 ± 0.14	0.15 ± 0.04	0.29 ± 0.17
ANOVA (df 2, 42)				
F ratio	1.28	1.03	1.16	0.33
P value	0.29	0.36	0.33	0.72

ponds, and thus provides some indication of stability among different community states. However, these data alone cannot be used to definitively test stability, and experimental perturbations of species densities would be required to unequivocally discern multiple stable equilibria (25).

The results from this long-term experiment show that the deterministic processes inherent in niche-based theories of community assembly predominate in lower-productivity systems, whereas stochastic processes due to differential colonization history and priority effects predominate in higher-productivity systems. The key mechanism underlying this relationship was that the realized pool of species that could potentially live in lower-productivity sites was more or less nested within the pool of species that could potentially live in higher-productivity sites. Although the generality of this specific mechanism in other ecosystems remains to be elucidated, I would argue that it is likely to be a rather general phenomenon. Species that are tolerant of harsh conditions such as low productivity can often be found on occasion in adjacent, more benign habitat types, and indeed can flourish in those habitats in the absence of interspecific interactions (26). For example, a large number of experiments have manipulated productivity in grassland ecosystems, and these experiments often show that lower-productivity treatments have a smaller, nested, pool of plant species that can potentially exist in those treatments relative to higher-productivity treatments that have a larger species pool (17). Similar patterns of nestedness along a productivity gradient seem to be evident in large-scale surveys of freshwater fish in lakes (10).

Some of the best evidence for stochastic processes underlying community assembly comes

from relatively productive environments such as tropical rainforests and coral reefs (13, 18, 27, 28), whereas deterministic processes may play a stronger role in less productive temperate forests (29, 30), grasslands (31, 32), and marine intertidal habitats (33). As such, the results from this experimental study are likely to be general, suggesting that the relative importance of stochasticity increases with increasing productivity, and providing a likely mechanism for the positive relationship between increasing productivity (and lower latitude) and increasing biodiversity at large spatial scales, even in the absence of systematic variation in habitat heterogeneity or biogeographic constraints.

References and Notes

- G. E. Hutchinson, *Am. Nat.* **93**, 145 (1959).
- J. H. Brown, *Am. Zool.* **27**, 877 (1981).
- M. L. Rosenzweig, *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
- R. Field *et al.*, *J. Biogeogr.* **36**, 132 (2009).
- G. G. Mittelbach *et al.*, *Ecology* **82**, 2381 (2001).
- J. M. Chase, M. A. Leibold, *Nature* **416**, 427 (2002).
- A. Bonn, D. Storch, K. J. Gaston, *Proc. Biol. Sci.* **271**, 1685 (2004).
- S. Harrison, K. F. Davies, H. D. Stafford, J. H. Viers, *J. Ecol.* **94**, 110 (2006).
- Y. Bai *et al.*, *J. Appl. Ecol.* **44**, 1023 (2007).
- T. Gardezi, A. Gonzalez, *Am. Nat.* **171**, 800 (2008).
- K. L. Evans, S. E. Newson, D. Storch, J. J. D. Greenwood, K. J. Gaston, *J. Anim. Ecol.* **77**, 395 (2008).
- K. S. He, J. Zhang, *Ecol. Inform.* **4**, 93 (2009).
- R. C. Condit *et al.*, *Science* **295**, 666 (2002).
- H. Tuomisto, K. Ruokolainen, M. Yli-Halla, *Science* **299**, 241 (2003).
- M. Dornelas, S. R. Connolly, T. P. Hughes, *Nature* **440**, 80 (2006).
- J. M. Chase, *Oecologia* **136**, 489 (2003).
- D. R. Chalcraft *et al.*, *Ecology* **89**, 2165 (2008).
- S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
- T. Fukami, *Popul. Ecol.* **46**, 137 (2004).
- C. F. Steiner, M. A. Leibold, *Ecology* **85**, 107 (2004).
- See supporting material on Science Online.
- J. M. Chase, M. A. Leibold, *Ecological Niches: Linking Classical and Contemporary Approaches* (Univ. of Chicago Press, Chicago, 2003).
- R. Ejrnæs, H. H. Bruun, B. J. Graae, *Ecology* **87**, 1225 (2006).
- J. M. Chase, *Oikos* **101**, 187 (2003).
- M. Scheffer, *Critical Transitions in Nature and Society* (Princeton Univ. Press, Princeton, NJ, 2009).
- J. P. Grime, *Plant Strategies, Vegetation Processes, and Ecosystem Properties* (Wiley, Chichester, UK, ed. 2, 2001).
- A. M. Latimer, J. A. Silander Jr., R. M. Cowling, *Science* **309**, 1722 (2005).
- I. Volkov, J. R. Banavar, S. P. Hubbell, A. Maritan, *Nature* **450**, 45 (2007).
- J. S. Clark, J. S. McLachlan, *Nature* **423**, 635 (2003).
- B. Gilbert, M. J. Lechowicz, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7651 (2004).
- J. Fargione, C. S. Brown, D. Tilman, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8916 (2003).
- J. M. Levine, J. HilleRisLambers, *Nature* **461**, 254 (2009).
- J. T. Wootton, *Nature* **433**, 309 (2005).
- I thank T. Knight, E. Marnocha, and C. Goss for help setting up the experiment; R. Shulman, B. Biro, J. Mihajlevic, P. Hanley, A. Strauss, and dozens of Tyson researchers for help sampling organisms; Tyson staff for logistical support; and T. Knight, K. Smith, R. Ricklefs, J. Losos, the reviewers, and my lab group for discussions and comments. Supported by Washington University and by NSF grants DEB-0241080 and 0816113.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1187820/DC1
Materials and Methods

SOM Text

Tables S1 and S2

References

2 February 2010; accepted 26 April 2010

Published online 27 May 2010;

10.1126/science.1187820

Include this information when citing this paper.

Exploitation of the Intestinal Microflora by the Parasitic Nematode *Trichuris muris*

K. S. Hayes, A. J. Bancroft, M. Goldrick, C. Portsmouth, I. S. Roberts,* R. K. Grencis*

The inhabitants of the mammalian gut are not always relatively benign commensal bacteria but may also include larger and more parasitic organisms, such as worms and protozoa. At some level, all these organisms are capable of interacting with each other. We found that successful establishment of the chronically infecting parasitic nematode *Trichuris muris* in the large intestine of mice is dependent on microflora and coincident with modulation of the host immune response. By reducing the number of bacteria in the host animal, we significantly reduced the number of hatched *T. muris* eggs. Critical interactions between bacteria (microflora) and parasites (macrofauna) introduced a new dynamic to the intestinal niche, which has fundamental implications for our current concepts of intestinal homeostasis and regulation of immunity.

The mammalian gut contains around 10^{13} bacteria (1), the majority of which belong to the phyla Bacteroidetes or Firmicutes (1, 2). Coevolution with these microbes has driven the functional morphology and immune function

of the gastrointestinal tract (3–5). Without microbes, aberrant physiology develops together with problems in host defense. Both can be rectified upon reintroduction of bacteria (6). Additionally, childhood exposure to microbes can

direct the maturing immune system to develop a tolerance to environmental antigens, the so-called “hygiene hypothesis” (7). More recently this concept has been extended to include “macrofauna” of the gut, such as helminth parasites (8). A helminth-driven T_H2 and regulatory helper T cell response has evolved to counter infection and repair the damage that these parasites cause (9). Dysregulation of these immune responses leads to prolonged infection and disease. Indeed, helminths have been found to be a major force underlying the evolution and selection of interleukin genes (10). Thus, gut commensal bacteria and gastrointestinal-dwelling helminths have lived in close association throughout evolution. Relationships between bacteria and metazoa have already been documented, such as filarial worms and the endosymbiont *Wolbachia* (11); however, a functional nonendosymbiotic relationship between prokaryotes and parasitic metazoa within the infected

Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK.

*To whom correspondence should be addressed. E-mail: richard.k.grencis@manchester.ac.uk (R.K.G.); ian.s.roberts@manchester.ac.uk (I.S.R.)

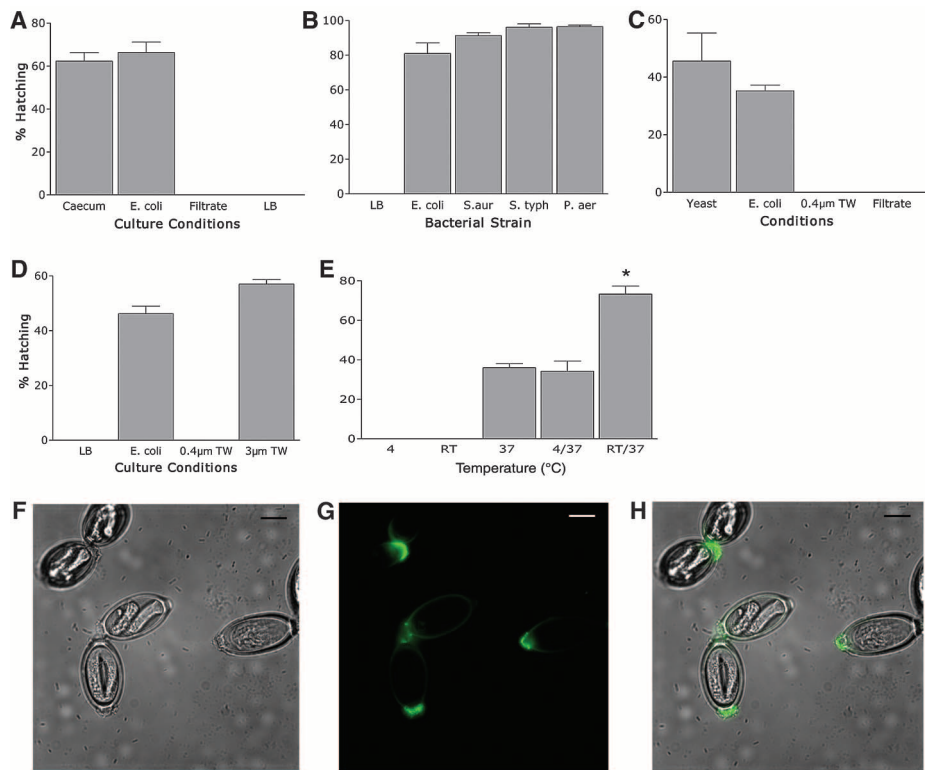


Fig. 1. *T. muris* eggs are induced to hatch in vitro by contact with bacteria. (A) Hatching of *T. muris* eggs for 2 hours at 37°C incubated with 5-cm sections of mouse cecum, *E. coli* bacterial suspension, or 0.22-µm filtered (overnight) bacterial suspension. Luria-Bertani (LB) broth was used as negative control. (B) *T. muris* eggs were cultured with four strains of bacteria (*E. coli*, *S. aureus*, *S. typhimurium*, or *P. aeruginosa*) for 2 hours at 37°C. (C) *T. muris* eggs incubated with *Saccharomyces cerevisiae*, with or without 0.4-µm transwell or 0.22-µm filtered *S. cerevisiae* cultures for 2 hours at 37°C. (D) Hatching of *T. muris* eggs in *E. coli* bacterial suspension or with 0.4-µm or 3-µm transwells. (E) *T. muris* eggs incubated for 2 hours with *E. coli* at 4°C, room temperature (RT), or 37°C. Eggs that did not hatch at 4°C or room temperature were then incubated for a further 2 hours at 37°C (4/37 and RT/37, respectively). (F to H) *T. muris* eggs were cultured at 37°C for 1 hour with GFP-expressing bacteria, then washed in phosphate-buffered saline before further incubation for 1 hour at 37°C. (F) Light-field image; (G) fluorescence image; (H) combined image. Scale bar, 10 µm. All figures show means ± SEM, **P* < 0.01, using analysis of variance (ANOVA).

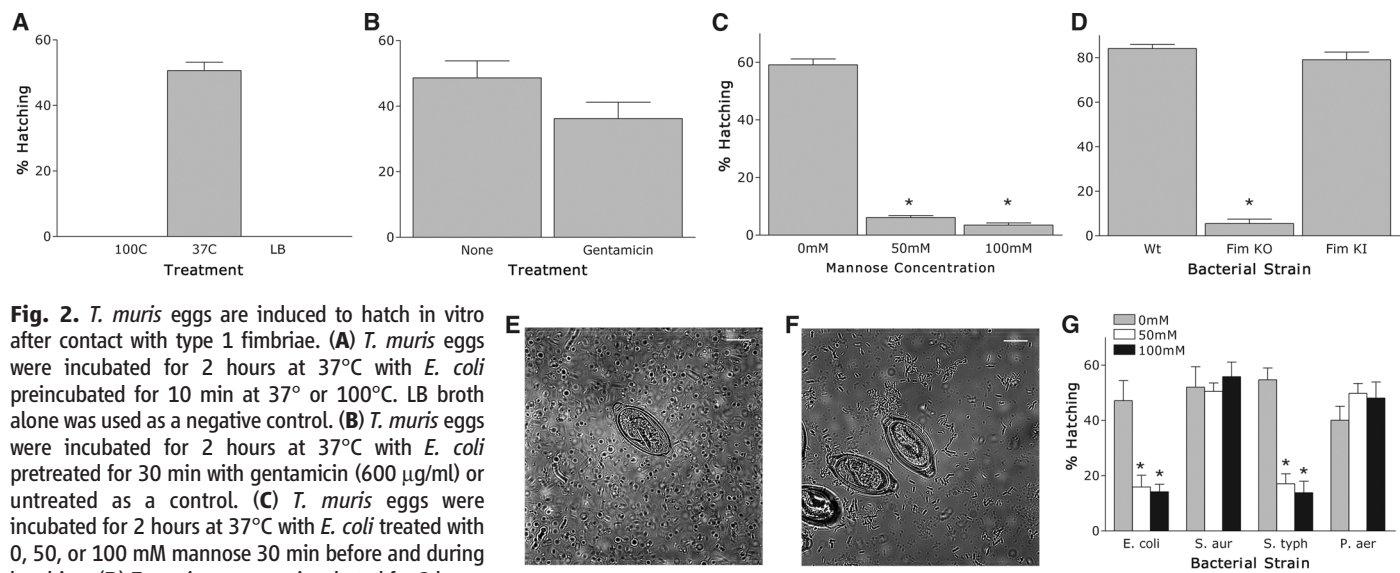


Fig. 2. *T. muris* eggs are induced to hatch in vitro after contact with type 1 fimbriae. (A) *T. muris* eggs were incubated for 2 hours at 37°C with *E. coli* preincubated for 10 min at 37° or 100°C. LB broth alone was used as a negative control. (B) *T. muris* eggs were incubated for 2 hours at 37°C with *E. coli* pretreated for 30 min with gentamicin (600 µg/ml) or untreated as a control. (C) *T. muris* eggs were incubated for 2 hours at 37°C with *E. coli* treated with 0, 50, or 100 mM mannose 30 min before and during hatching. (D) *T. muris* eggs were incubated for 2 hours at 37°C with wild-type *E. coli* (strain PK1162), FimKO *E. coli*, or FimKI *E. coli*. (E and F) *T. muris* eggs were incubated with FimKO (E) or FimKI (F) *E. coli* for 1 hour at 37°C. Scale bar, 10 µm. (G) *T. muris* eggs after 2 hours of incubation

at 37°C with *E. coli*, *S. aureus*, *S. typhimurium*, or *P. aeruginosa* plus 0, 50, or 100 mM mannose before and during hatching. All figures show means ± SEM, **P* < 0.001 using ANOVA.

host remains to be defined. This complex intestinal ecology will have major implications for the immunoregulatory mechanisms of gut inflammation and autoimmune disease (12).

Trichuris is a genus comprising more than 50 species of whipworm, an extremely prevalent and successful group of intestinal-dwelling nematode parasites infecting many diverse mammalian hosts, with *T. trichiura* estimated to infect almost 1 billion people (13). All *Trichuris* species inhabit the large intestine (cecum and colon). Infection proceeds upon ingestion of embryonated eggs from the external environment. Upon hatching, the larvae emerge from polar egg opercula and establish infection within the epithelium of the crypts of Lieberkühn of the cecum and colon. Following the characteristic four molts, dioecious adult parasites develop to patency (at a rate dependent on the host), mate, and release unembryonated eggs into the environment via the feces. Here, we investigated a bacterially driven mechanism of hatching of mouse whipworm, *T. muris*, that facilitates infection of the mammalian host and subsequent immune response.

T. muris eggs were induced to hatch in vitro when incubated for at least 30 min with explants of mouse cecum containing substantial numbers of bacteria at 37°C (Fig. 1A). To define the role of bacteria in this process, we incubated eggs in a culture of *Escherichia coli*, a common gut commensal. In the presence of *E. coli*, hatching was observed at a similar level to that seen with gut explants; 0.4-µm filtration to remove *E. coli* from cultures prevented hatching and showed that a structural component of the bacteria, not a secreted molecule, was responsible for the hatching (Fig. 1A). Further analysis confirmed that a variety of microorganisms (five strains of bacteria and one of yeast) could induce efficient hatching over 2 hours to levels comparable with that found with

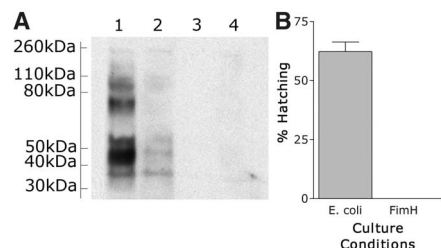


Fig. 3. (A) Far Western blot analysis of purified FimH binding to surface egg proteins. *T. muris* eggs are induced to hatch in vitro after contact with FimH adhesion on type 1 fimbriae. Egg proteins were subjected to SDS–polyacrylamide gel electrophoresis under reducing conditions and electrophoretically transferred onto nitrocellulose. The migration positions of protein standards are indicated on the left. Lane 1, purified FimH and secondary antibody; lane 2, purified FimH pre-incubated with mannose to specifically block binding; lane 3, purified FimH only (no secondary); lane 4, secondary only. **(B)** Purified FimH does not by itself cause hatching of *T. muris* eggs.

gut explants over 18 hours (Fig. 1, B and C). By using transwells of different sizes, it was possible to confirm that direct contact between the bacteria and the eggs was required for hatching (Fig. 1, C and D). Bacterially promoted hatching only occurred at 37°C, suggesting that temperature is also a hatching cue (Fig. 1E), presumably to prevent hatching in the external environment where *T. muris* eggs embryonate. To locate the site of interaction between the bacteria and parasite eggs, we incubated the green fluorescent protein (GFP)–expressing *E. coli* strain PK1162 (14) with embryonated eggs. Bacteria clearly cluster around the opercula at the poles of the eggs (Fig. 1, F to H) where the worms emerge (movie S1).

Structural disruption of *E. coli* (boiling, 10 min) prevented hatching (Fig. 2A), although after bacteriostatic antibiotic treatment (gentamicin), *E. coli* were still able to induce hatching (Fig. 2B) but were not viable. Taken together, the data support a role for the intact bacterial surface as a critical component of the hatching process.

Type 1 fimbriae facilitate mannose-sensitive adherence of *E. coli* to cells and mucosal surfaces (15). They are encoded by the *fim* gene cluster and consist of a major structural subunit (FimA) and several minor components including the adhesin FimH located at the fimbrial tip that recognizes terminally located D-mannose moieties on cell-bound and secreted glycoproteins (16, 17). To elucidate whether type 1 fimbriae played a role in hatching, we investigated the effect of exogenous mannose on hatching. The addition of mannose significantly inhibited hatching without affecting the numbers or viability of bacteria (Fig. 2C and fig. S1). Likewise, strain AAEC072A (FimKO), which lacks the *fim* gene cluster and does not express type 1 fimbriae (18) (fig. S2), did not promote hatching (Fig. 2D), although viability of these bacteria was comparable to that of wild-type Fim⁺ *E. coli* strain PK1162 (fig. S3). In contrast, mannose-

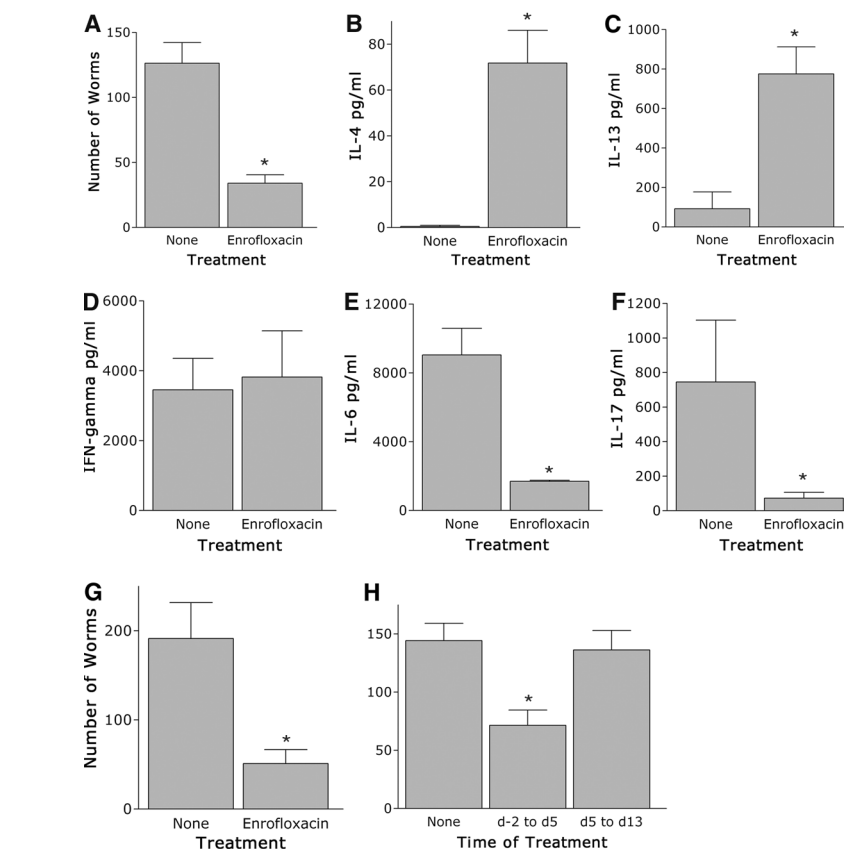


Fig. 4. *T. muris* eggs are induced to hatch in vivo by the presence of bacteria. **(A)** Worm burden at day 18 p.i. in AKR mice treated with enrofloxacin (Baytril) from day –2 p.i. to day 18 p.i. Untreated AKR mice were used as controls. **(B)** IL-4 secretion by antigen restimulated mesenteric lymph node cells from enrofloxacin-treated and untreated control AKR mice at day 18 p.i. assessed by cytokine bead array. **(C)** IL-13 secretion. **(D)** IFN-γ secretion. **(E)** IL-6 secretion. **(F)** IL-17 secretion. **(G)** Worm burden at day 18 p.i. in SCID mice treated with enrofloxacin from day –2 p.i. to day 18 p.i. Untreated SCID mice were used as controls. **(H)** Worm burden at day 14 p.i. in AKR mice treated with enrofloxacin at the beginning only (day –5 to day 5) or end only (day 5 to day 13) of infection. Untreated AKR mice used as controls. All figures show means ± SEM, **P* < 0.01 by ANOVA.

sensitive hatching was seen with strain AAEC072A harboring plasmid pLB254 (FimKI) encoding the cloned *fim* gene cluster (Fig. 2D and fig. S2). In contrast to the strains expressing type 1 fimbriae that associate with the opercula at the poles (Fig. 1, F to H, and Fig. 2, E and F), strain AAEC072A failed to adhere to the eggs (Fig. 2, E and F), confirming a role for type 1 fimbriae in mediating interactions between the parasite eggs and the bacterium. The observation that *Salmonella typhimurium* also exhibited mannose-sensitive hatching similar to that seen with *E. coli* (Fig. 2G) indicates that bacterially induced hatching is partly mediated by type 1 fimbriae. However, the Gram-negative bacterium *Pseudomonas aeruginosa* induced hatching in the presence of mannose and does not express type 1 fimbriae. Similarly, the ability of the Gram-positive bacterium *Staphylococcus aureus* to promote hatching indicates that other mechanisms also exist.

His-tagged FimH was purified (fig. S4) and used in a far Western blot analysis of solubilized surface egg proteins. Five to seven surface proteins were detected that bound to purified FimH,

with the most intense signal generated from a protein of 45 kD (Fig. 3A). Mass spectrophotometry of these different bands and data mining of an expressed sequence tag database (NEMBASE) have identified putative proteins of *T. muris* and other parasitic nematodes, although further categorization will depend on sequencing of the *T. muris* genome. Purified FimH itself did not cause hatching of *T. muris* eggs (Fig. 3B), suggesting that some structural conformation, perhaps cross-linking by multiple FimH adhesins, is required. In addition, it is known that type 1 fimbriae mediate shear-dependent adhesion typified by weak binding under conditions of low flow that strengthens as the shear forces increase (17). Thus, it is perhaps not surprising that FimH on the bacterial surface behaves differently from the purified molecule. It is known that interaction between FimH and uroplakin molecules on the cell surface of uroepithelial umbrella cells or β₁ integrins on fibroblast cells is known to induce a signal transduction cascade (19). It is possible that the interaction between FimH and a mannosylated receptor on the surface of the *T. muris* eggs stimulates a

signal transduction cascade that leads to the emergence of the worm out of the egg.

Although the data clearly show that hatching in vitro requires an interaction between bacteria and the egg via type 1 fimbriae, it is unclear whether this is also important in vivo. Thus, we treated AKR mice with enrofloxacin from 2 days before infection throughout the course of infection to ascertain whether there was any effect on worm establishment. At day 21 post-infection (p.i.), the worm burden was significantly ($P < 0.001$) decreased in animals that had been treated with antibiotics (Fig. 4A). Interestingly, these animals made a stronger T_H2 response to infection, with increased levels of interleukin-4 (IL-4) (Fig. 4B) and IL-13 (Fig. 4C). Interferon- γ (IFN- γ) secretion was unchanged (Fig. 4D). Parasite-specific immunoglobulin G1 (IgG1) antibody production was increased in antibiotic-treated animals, reflecting the increased T_H2 response, whereas IFN- γ -controlled IgG2a levels were unaffected, reflecting the IFN- γ response (fig. S5). IL-17 has been found to be related to bacterial presence in the small intestine (20) and is significantly reduced in animals treated with antibiotics. Similarly, in our experiments, IL-17 expression was reduced in antibiotic-treated animals (Fig. 4F), as was IL-6, which drives IL-17 production (Fig. 4E).

To rule out a major role for acquired immunity in worm reduction, we infected severe combined immunodeficient (SCID) mice and treated them with enrofloxacin throughout. At day 18 p.i., worm burdens were reduced in antibiotic-treated mice (Fig. 4G), demonstrating that it is not the increased T_H2 response that is solely responsible for the worm expulsion observed. A comparison of antibiotic treatment regimes in susceptible mice (treatment from day -5 to day +5 p.i. or day +7 to day 13 p.i.) clearly showed that depleting the microflora affected the establishment and not the subsequent survival of the worms within the intestine at day 14 p.i. (Fig. 4H). All animals treated with antibiotics had decreased aerobic and anaerobic flora as measured by growth on LB agar (fig. S6).

The paradigm that the adaptive immune system has evolved to control microbes has been modified to include the concept that the immune system is in fact controlled by microorganisms. Our results constitute evidence to suggest that we must extend this paradigm again to include "macrofauna," the metazoan parasites such as *Trichuris*. We can clearly see a critical relationship between bacteria and *T. muris* that has a marked effect on the host immune response to this parasite. We propose that temperature, together with an increase in bacterial load at these sites, provides optimum conditions for efficient hatching of *Trichuris* species eggs. Additionally, controlled damage of the intestinal epithelium provides an enhanced potential route of access for microflora to the adaptive immune system. This has subsequent beneficial consequences for the host, in that antibacterial T_H17 responses and T_H1 responses are generated that are known to

control pathogenic bacterial pathogens (20, 21). The induction of T_H17 and T_H1 responses in the gut is closely associated with the induction of regulatory T cell subsets (22). These are known to have beneficial effects for survival of long-lived chronic helminth infections in the absence of overt pathology and indeed are central to the tenet of the modified hygiene hypothesis (8). It is clear that in nature, coevolution of microflora of the gut and host should not be considered in the absence of the influence of macrofauna such as gastrointestinal-dwelling nematode parasites.

References and Notes

1. R. E. Ley *et al.*, *Science* **320**, 1647 (2008); published online 22 May 2008 (10.1126/science.1155725).
2. P. J. Turnbaugh *et al.*, *Nature* **449**, 804 (2007).
3. L. V. Hooper, J. I. Gordon, *Science* **292**, 1115 (2001).
4. F. Bäckhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, J. I. Gordon, *Science* **307**, 1915 (2005).
5. S. M. Edelman, D. L. Kasper, *Curr. Opin. Gastro.* **24**, 720 (2008).
6. L. V. Hooper, T. S. Stappenbeck, C. V. Hong, J. I. Gordon, *Nat. Immunol.* **4**, 269 (2003).
7. A. E. Wold, *Allergy* **53** (suppl.), 20 (1998).
8. M. Yazdanbakhsh, P. M. Matricardi, *Clin. Rev. Allergy Immunol.* **26**, 15 (2004).
9. R. M. Maizels, M. Yazdanbakhsh, *Nat. Rev. Immunol.* **3**, 733 (2003).
10. M. Fumagalli *et al.*, *J. Exp. Med.* **206**, 1395 (2009).
11. M. J. Taylor, C. Bandi, A. Hoerauf, *Adv. Parasitol.* **60**, 245 (2005).
12. V. Gaboriau-Routhiau *et al.*, *Immunity* **31**, 677 (2009).
13. *Prevention and Control of Schistosomiasis and Soil-Transmitted Helminthiasis*, WHO Tech. Rep. 912 (2002).
14. A. Reisner, J. A. Haagensen, M. A. Schembri, E. L. Zechner, S. Molin, *Mol. Microbiol.* **48**, 933 (2003).
15. I. B. Ofek, E. H. Beachey, *Infect. Immun.* **22**, 247 (1978).
16. D. S. Eto, T. A. Jones, J. L. Sundsbak, M. A. Mulvey, *PLoS Pathog.* **3**, e100 (2007).
17. J. Bouckaert *et al.*, *Mol. Microbiol.* **61**, 1556 (2006).
18. D. M. Breines, J. C. Burnham, *J. Antimicrob. Chemother.* **34**, 205 (1994).
19. P. Thumbikat *et al.*, *PLoS Pathog.* **5**, e1000415 (2009).
20. I. I. Ivanov *et al.*, *Cell Host Microbe* **4**, 337 (2008).
21. I. I. Ivanov *et al.*, *Cell* **139**, 485 (2009).
22. A. Chaudhry *et al.*, *Science* **326**, 986 (2009); published online 1 October 2009 (10.1126/science.1172702).
23. We thank L. Cliffe for numerous helpful discussions; J. Kott and R. Fernandez for help with microscopy; the staff in the Faculty of Life Sciences EM Facility for their assistance, particularly A. Mironov and the Wellcome Trust for equipment grant support to the EM Facility; the Faculty of Life Sciences Mass Spectrometry Unit and E. Keavill for help with protein identification; and P. Klemm for bacterial strains and plasmids. The Bioimaging Facility microscopes used in this study were purchased with grants from the UK Biotechnology and Biological Sciences Research Council, Wellcome, and the University of Manchester Strategic Fund. Accession numbers for proteins identified by mass spectrometry are Q9N5A6 CAEEL, TMP00496_1, TMP00827_1, TMP00007_1, and TMP00821_1 and can be found at www.nematodes.org/nembase4. Supported by Wellcome Trust grant 083620Z and UK Medical Research Council grant G0601205.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1391/DC1
Materials and Methods
Figs. S1 to S6
Movie S1
References

29 January 2010; accepted 30 April 2010
10.1126/science.1187703

Defective Cross-Presentation of Viral Antigens in GILT-Free Mice

Reshma Singh and Peter Cresswell*

Gamma-interferon-inducible lysosomal thiolreductase (GILT) promotes major histocompatibility complex (MHC) class II-restricted presentation of exogenous antigens containing disulfide bonds. Here, we show that GILT also facilitates MHC class I-restricted recognition of such antigens by $CD8^+$ T cells, or cross-presentation. GILT is essential for cross-presentation of a $CD8^+$ T cell epitope of glycoprotein B (gB) from herpes simplex virus 1 (HSV-1) but not for its presentation by infected cells. Initiation of the gB-specific $CD8^+$ T cell response during HSV-1 infection, or cross-priming, is highly GILT-dependent, as is initiation of the response to the envelope glycoproteins of influenza A virus. Efficient cross-presentation of disulfide-rich antigens requires a complex pathway involving GILT-mediated reduction, unfolding, and partial proteolysis, followed by translocation into the cytosol for proteasomal processing.

Cross-priming (1) is important for the development of specific $CD8^+$ T cell responses to viruses that do not directly infect antigen-presenting cells (APCs) (2). The critical APCs for cross-presentation are dendritic cells (DCs), which acquire antigens by phagocytosis of apoptotic and necrotic infected cells and migrate to secondary lymphoid organs to activate resident naïve $CD8^+$ T cells (3). Transfer of antigen from migratory DCs to resident $CD8\alpha^+$ DCs may be required (4, 5). The pathways that generate complexes of MHC class I molecules with peptides derived from internalized

antigens are not well understood. Occasionally the peptides are generated in the endocytic pathway and bind to recycling MHC class I molecules (6). However, the dominant mechanism involves translocation of the antigens into the cytosol, where proteasomal degradation generates peptides that

Department of Immunobiology, Howard Hughes Medical Institute, Yale University School of Medicine, 300 Cedar Street, Post Office Box 208011, New Haven, CT 06250-8011, USA.

*To whom correspondence should be addressed. E-mail: peter.cresswell@yale.edu

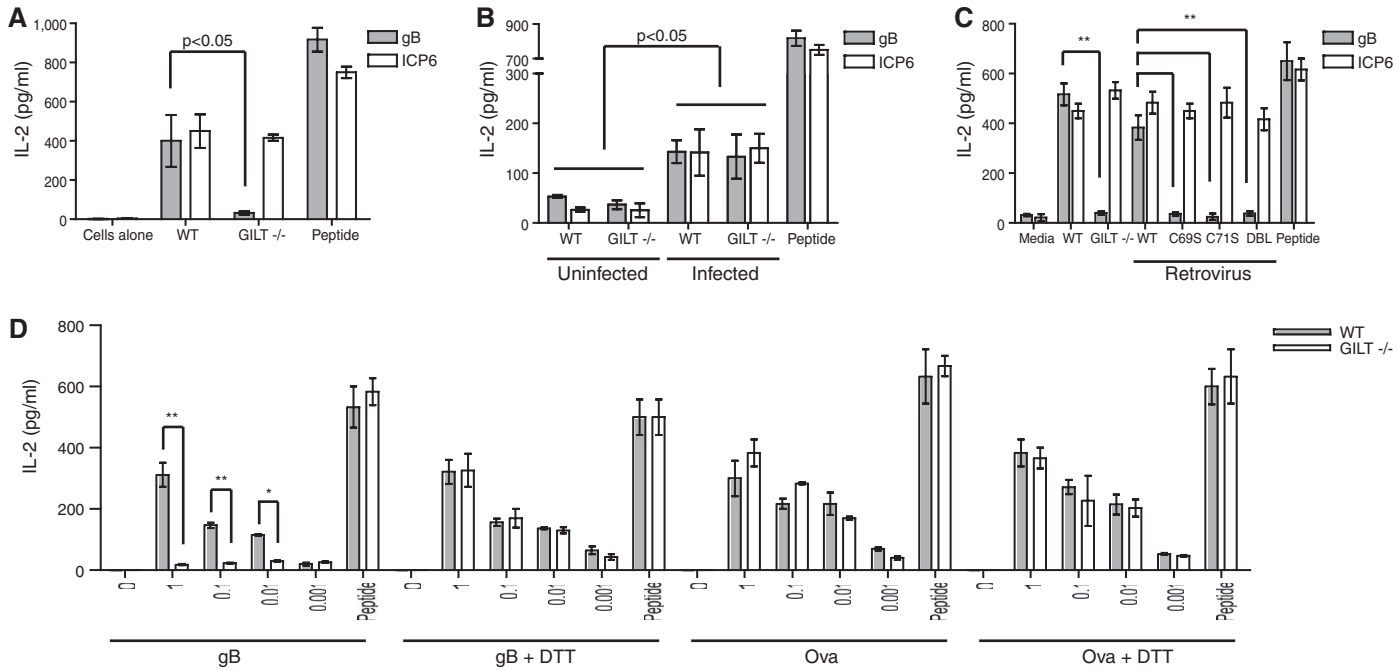


Fig. 1. Reduction by GILT is necessary for cross-presentation of gB₄₉₈₋₅₀₅. Interleukin-2 (IL-2) production, measured by enzyme-linked immunosorbent assay (ELISA), by K^b-restricted gB₄₉₈₋₅₀₅- and ICP₆₈₂₂₋₈₂₉-specific CD8⁺ T cell hybridomas in response to (A) wild-type (WT) and GILT-negative DCs co-cultured with HSV-1-infected HeLa cell debris, (B) WT and GILT-negative DCs directly infected by HSV-1, or (C) GILT-negative DCs retrovirally transduced with WT GILT or inactive single or

double cysteine mutants after co-culture with infected cell debris. (D) IL-2 production by gB₄₉₈₋₅₀₅- and ovalbumin-specific CD8⁺ T cell hybridomas in response to WT and GILT-negative DCs co-cultured with the indicated concentrations of recombinant soluble gB or ovalbumin (Ova), untreated or treated with dithiothreitol (DTT). Each of the panels shown is representative of three individual experiments. **P* < 0.05; ***P* < 0.01, calculated by Student's *t* test. Graphs show mean ± SEM.

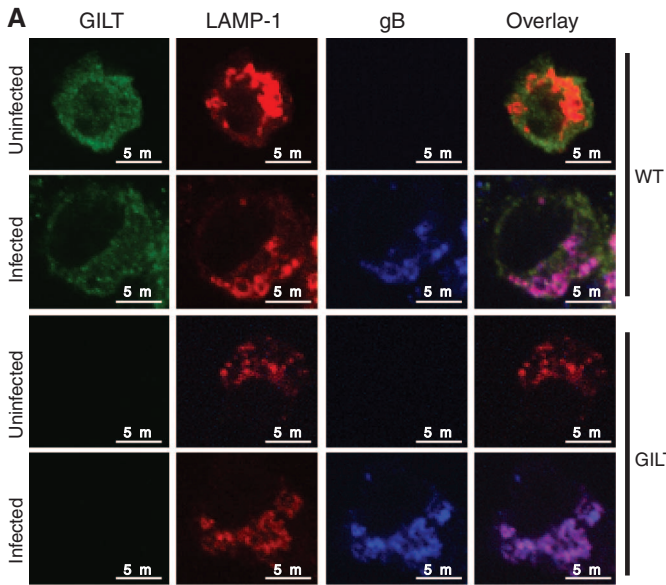
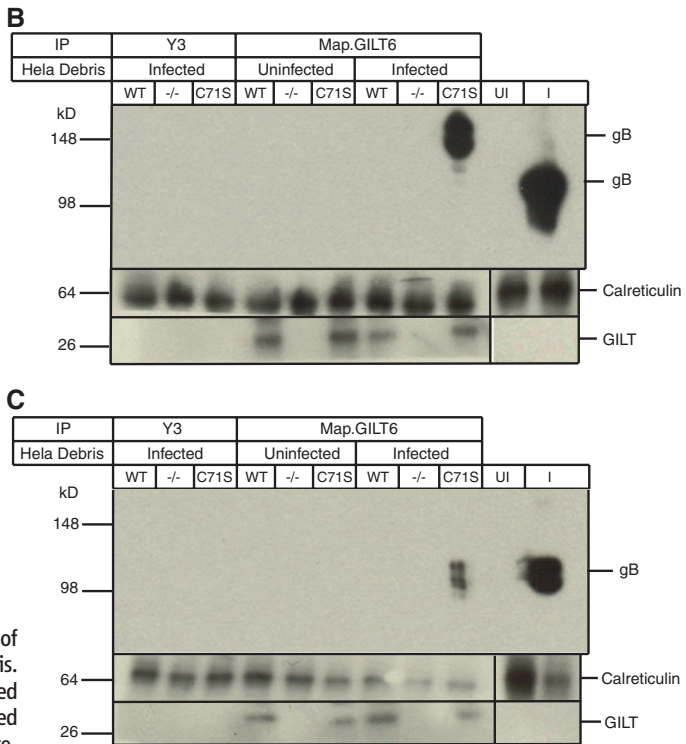


Fig. 2. GILT interacts with gB in DCs. (A) Visualization of intracellular location of GILT and gB WT and GILT-negative DCs that have taken up infected HeLa cell debris. WT and GILT-negative DCs were incubated with either uninfected or HSV-1-infected HeLa cell debris for 3 hours. Cells were then harvested, permeabilized, and stained for immunofluorescence. (B) WT DCs, GILT-negative DCs, or GILT-negative DCs reconstituted with the GILT C71S trapping mutant were incubated with infected HeLa cell debris for 3 hours before detergent solubilization and immunoprecipitation (IP) with a control antibody to H2-K^b (Y3) or a GILT monoclonal antibody (MaP.GILT6), nonreducing SDS-PAGE, and Western blotting. Top panel: Gel was probed with a gB-specific rabbit antiserum. Middle panel: The DC or HeLa cell lysates were probed with mouse or human antibodies to calreticulin as a loading control. Bottom panel: Lysates were probed with an antibody to GILT. GILT is present only in WT DCs and the GILT-negative DC samples reconstituted with the mutant. The first nine lanes are DC lysates. The final two lanes in the top panel are uninfected (UI) or infected (I) HeLa cell lysates. (C) Identical to (B) except that SDS-PAGE was performed under reducing conditions. Each experiment was done at least three times, and a representative experiment is shown.



are transported via the transporter associated with antigen processing (TAP) and bind to newly synthesized MHC class I molecules (7). The translocation mechanism may involve components of the endoplasmic reticulum-associated degradation (ERAD) machinery (8, 9).

Intact functional proteins can enter the DC cytosol after internalization (10–12), and recently we showed that luciferases can be unfolded in the endocytic pathway, translocated, and cytosolically refolded by the chaperone Hsp90 (12). The suggestion that translocation may require unfolding led us to investigate the role of GILT, a soluble enzyme expressed constitutively in APCs, in cross-presentation. GILT is the only known thiol reductase localized in lysosomes and phagosomes (13, 14), and we hypothesized that acidification combined with GILT-mediated reduction could

mediate the unfolding of internalized disulfide-containing antigens and facilitate their translocation into the cytosol.

Viral glycoproteins are often recognized by CD8⁺ T cells and are rich in disulfide bonds. We selected gB from HSV-1, which has a well-characterized MHC class I-restricted epitope (15), as a model antigen. In vitro cross-presentation assays were established with bone marrow-derived DCs from wild-type mice and mice lacking *Irf30*, the gene encoding GILT, and HSV-infected HeLa cells to provide apoptotic or necrotic bodies for antigen uptake (16). Using an H2-K^b-restricted CD8⁺ T cell hybridoma specific for gB₄₉₈₋₅₀₅, we found that cross-presentation of gB is indeed dependent on GILT expression (Fig. 1A). Cross-presentation of a second HSV-1 epitope, ICP6₈₂₂₋₈₂₉ from a viral ribonucleotide reduc-

tase, was GILT independent. gB contains five disulfide bonds (17), whereas ICP6 is a cytosolic protein and likely has none. Wild-type and GILT-negative DCs presented both the ICP6 and gB epitopes when directly infected by HSV-1 (Fig. 1B). To determine whether the enzymatic activity of GILT is required for cross-presentation, we introduced wild-type GILT or single or double cysteine active-site mutants into *Irf30*^{−/−} DCs. Only wild-type GILT restored cross-presentation (Fig. 1C). Cross-presentation of purified recombinant gB by DCs was also mediated by wild-type DCs but not by those lacking GILT (Fig. 1D). If the disulfide bonds in gB were first reduced, however (fig. S1), GILT-negative cells were able to cross-present the gB epitope (Fig. 1D). Both wild-type and GILT-negative DCs were capable of cross-presenting an ovalbumin epitope regardless of reduction (Fig. 1D).

A central question is whether cross-presentation depends on reduction of intact gB by GILT. Immunofluorescence analysis showed that GILT and gB were both present in the same intracellular compartment as lysosomal-associated membrane protein 1 (LAMP-1), a lysosomal and phagosomal marker, in DCs incubated with necrotic infected HeLa cells (Fig. 2A). To demonstrate that GILT mediates gB reduction, we used a GILT trapping mutant with a mutation in the second cysteine of the CXXC active site, which leads to accumulation of disulfide-linked enzyme-substrate complexes because substrate release is blocked (14, 18). When necrotic infected HeLa cells were incubated with DCs expressing the trapping mutant, a gB-GILT mixed disulfide was readily detectable (Fig. 2B). Under reducing conditions, the GILT-associated gB had the same mobility in SDS-polyacrylamide gel electrophoresis (SDS-PAGE) as in the HeLa cells (Fig. 2C). The doublet likely results from differential glycosylation. These data provide evidence that GILT directly reduces disulfide bonds in the intact glycoprotein.

Vesicular acidification is usually required for MHC class II presentation and may be required for cross-presentation (19–21). Blocking acidification with bafilomycin or concanamycin B abrogated cross-presentation of the gB₄₉₈₋₅₀₅ epitope (Fig. 3A). GILT has an acidic pH optimum, but neutralization could also inhibit essential lysosomal proteolysis. We examined the effects of pepstatin A, which mainly inhibits cathepsin D, an aspartyl protease, and leupeptin, an inhibitor of cysteine proteases, including cathepsin B. Both blocked gB cross-presentation, suggesting that multiple proteases are required (Fig. 3, B and C). Furthermore, when we examined by immunofluorescence microscopy the turnover of gB within wild-type and GILT-negative DCs incubated with necrotic infected cells, gB expression decreased much more rapidly in the wild-type DCs (Fig. 3D). The data suggest that an interplay between GILT-mediated reduction and degradation by several proteases generates gB fragments that are then cross-presented.

Proteolysis in the phagosome could give rise to gB₄₉₈₋₅₀₅ that binds directly to K^b molecules or

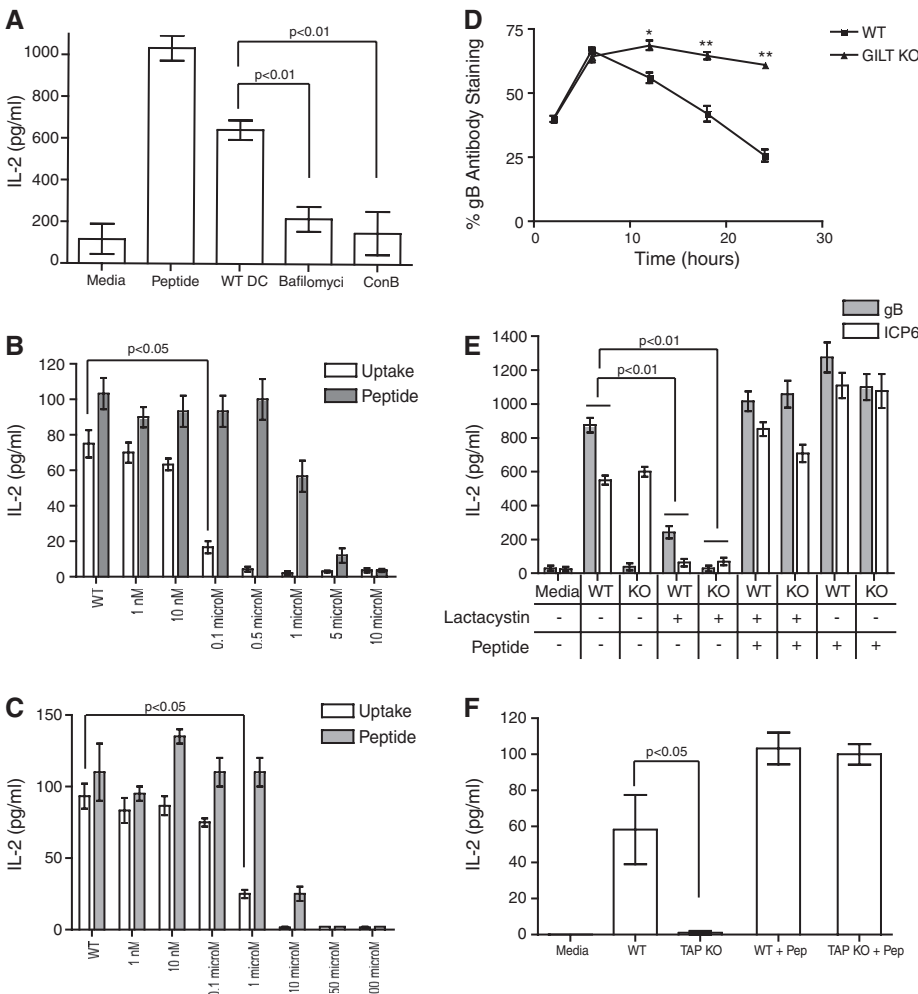


Fig. 3. GILT-dependent cross-presentation of gB requires lysosomal and proteasomal processing and is TAP dependent. (A) IL-2 production, measured by ELISA, by K^b-restricted gB₄₉₈₋₅₀₅-specific CD8⁺ T cell hybridoma in response to WT DCs treated with bafilomycin or concanamycin B (ConB) before HeLa cell uptake. IL-2 production by gB₄₉₈₋₅₀₅-specific CD8⁺ T cell hybridomas in response to WT DCs treated with (B) pepstatin A or (C) leupeptin. (D) Kinetics of gB degradation, determined by immunofluorescence, in WT or GILT-negative DCs incubated with infected HeLa cell debris. A total of 1000 DCs were counted per time point and analyzed by staining with antibody to gB. (E) IL-2 production by gB₄₉₈₋₅₀₅- and ICP6₈₂₂₋₈₂₉-specific CD8⁺ T cell hybridomas to WT or GILT-negative DCs treated with lactacystin. (F) IL-2 production by gB₄₉₈₋₅₀₅-specific CD8⁺ T cell hybridomas to WT or *Tap1*^{−/−} DCs. A representative example of three individual experiments is shown for each panel. *P < 0.05; **P < 0.01, calculated by Student's *t* test. Graphs show mean ± SEM.

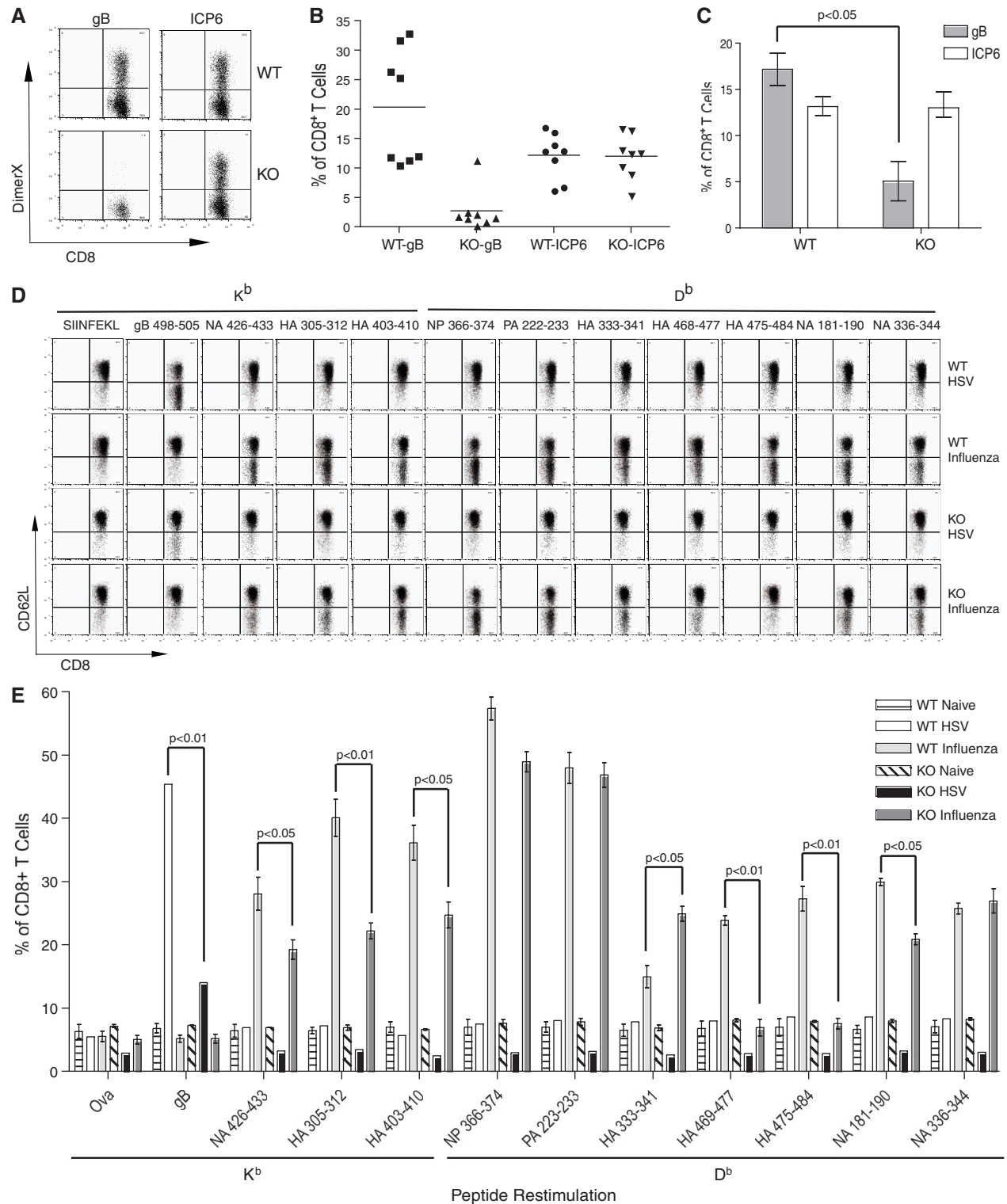


Fig. 4. GILT-dependent cross-priming to gB and influenza A virus glycoproteins in vivo. **(A)** Flow cytometric analysis of CD8⁺ T cells specific for gB₄₉₈₋₅₀₅ from WT or *Irf30*^{-/-} mice infected with HSV-1, detected with gB₄₉₈₋₅₀₅- or ICP6₈₂₂₋₈₂₉-loaded DimerX:K^b-Ig fusion protein. Draining LNs were examined 6 days after infection. A representative dot plot is shown. **(B)** The percentage of gB₄₉₈₋₅₀₅- and ICP6₈₂₂₋₈₂₉-specific CD8⁺ T cells from HSV-1-infected WT and *Irf30*^{-/-} mice. A representative experiment of three individual experiments is shown. **(C)** As in (B), showing the average of three independent experiments. Graph shows mean ±

SEM. **(D)** Flow cytometric analysis of recall CD8⁺ T cell responses isolated from HSV-1- or influenza A-infected mice and restimulated with WT DCs pulsed with the indicated peptides. Cells were cultured for 2 days with APCs loaded with each peptide before flow cytometric analysis for activation assessed by the down-regulation of CD62L. A representative dot plot of a WT mouse and a *Irf30*^{-/-} mouse infected with HSV or influenza A is shown. **(E)** As in (D), but showing the average of three independent experiments. Graphs show mean ± SEM. *P* values were calculated by Student's *t* test.

result in gB fragments that are translocated into the cytoplasm. To determine whether cytosolic access is required, we examined the roles of TAP and proteasomes in gB cross-presentation. When DCs from *Tap*^{−/−} mice were incubated with necrotic infected cells, gB cross-presentation was completely eliminated (Fig. 3F). In addition, cross-presentation of gB, as well as ICP6, was inhibited by lactacystin, indicating dependence on proteasomal processing (Fig. 3E). Cross-presentation thus depends on cytosolic processing of gB fragments generated in the phagosome by GILT-mediated reduction and cathepsin-mediated proteolysis.

A requirement for GILT in the induction of the CD8⁺ T cell response to gB₄₉₈₋₅₀₅ during an infection would argue that cross-priming is important for the in vivo anti-HSV-1 immune response. Wild-type and *Irf30*^{−/−} mice were infected with HSV-1, and the draining lymph nodes (LNs) were examined for the induction of K^b-gB₄₉₈₋₅₀₅-specific and K^b-ICP6₈₂₂₋₈₂₉-specific CD8⁺ T cells. Although mice lacking GILT generated the same average percentage of ICP6₈₂₂₋₈₂₉-specific CD8⁺ T cells when infected with HSV-1 as wild-type mice, the number of gB₄₉₈₋₅₀₅-specific CD8⁺ T cells was significantly reduced (Fig. 4, A to C). There was no difference in the survival of the infected mice. Responses to GILT-independent epitopes such as ICP6₈₂₂₋₈₂₉ may make up for any deficiency.

To determine whether GILT-dependent cross-presentation is a more general phenomenon, we examined the CD8⁺ T cell response of mice infected with the PR8 strain of influenza A virus. LN cells from naïve and infected mice were restimulated with wild-type DCs pulsed with peptides that correspond to a variety of H2-K^b- and H2-D^b-restricted epitopes from hemagglutinin (HA), neuraminidase (NA), polymerase (PA), and nucleoprotein (NP) (www.immuneepitope.com/home.do) (22). HA and NA contain six and eight disulfide bonds, respectively, whereas PA and NP have none (23–25). A similar percentage of wild-type and GILT-negative CD8⁺ T cells responded to D^b-restricted PA and NP epitopes upon restimulation (Fig. 4, D and E). In contrast, the responses of CD8⁺ T cells from mice lacking GILT were significantly reduced for four out of five of the HA epitopes and for two out of three of the NA epitopes. The two HA epitopes to which almost no CD8⁺ T cells develop in the *Irf30*^{−/−} mice contain or are immediately adjacent to a cysteine (C480) involved in a disulfide bond (C21-C480). For both HA and NA, one epitope is GILT independent, strongly arguing against the possibility that any GILT requirement reflects GILT-dependent MHC class II-restricted responses that mediate CD4⁺ T cell help (26). Although the epitope specificity of the CD4⁺ T cells in the *Irf30*^{−/−} mice may be different from that of wild-type mice, the total numbers of CD4⁺ T cells that they generate during a viral immune response are similar (fig. S2), as are the numbers of CD4⁺ T cells in the spleens of uninfected wild-type and *Irf30*^{−/−} mice. The data show that GILT-dependent cross-presentation is not restricted to gB, and that cross-priming is important

in the CD8⁺ T cell response to influenza virus. The residual CD8⁺ T cell responses observed to gB and the HA and NA epitopes by *Irf30*^{−/−} animals may reflect priming by directly infected APCs.

The only known function of GILT is to reduce disulfide bonds, and we have shown that GILT is essential for cross-presentation of many peptides from disulfide-containing proteins. We suggest that reduction in the acidic environment of the phagosome facilitates partial proteolysis into fragments that are translocated into the cytosol where they are further degraded by the proteasome to generate peptides. These are transported by TAP and bind in a conventional manner, possibly after amino-terminal trimming (27), to MHC class I molecules. This latter step is likely to occur in the ER, but could occur in phagosomes that have recruited ER membrane components, although this issue remains contentious (28, 29).

For gB, the inability to cross-present is reflected in a reduction in K^b-gB₄₉₈₋₅₀₅-specific CD8⁺ T cells in vivo, indicating the importance of cross-priming in CD8⁺ T cell responses to HSV-1 infection. The similar reduction in HA- and NA-specific CD8⁺ T cells suggests that cross-priming is also important during influenza A infection. The role played by GILT in cross-priming, combined with its established involvement in MHC class II-restricted CD4⁺ T cell responses (30), indicates the importance of the enzyme in the immune system. This may have implications for vaccine design and approaches to tumor immunotherapy that involve peptide-based vaccines, in that linear peptides may not be the optimal vehicles for the expression of GILT-dependent epitopes, and for autoimmunity to self-antigens that contain multiple disulfide bonds.

References and Notes

1. M. J. Bevan, *J. Exp. Med.* **143**, 1283 (1976).
2. L. J. Sigal, S. Crotty, R. Andino, K. L. Rock, *Nature* **398**, 77 (1999).

3. K. L. Rock, L. Shen, *Immunol. Rev.* **207**, 166 (2005).
4. R. S. Allan *et al.*, *Immunity* **25**, 153 (2006).
5. S. Bedoui *et al.*, *Nat. Immunol.* **10**, 488 (2009).
6. R. Belizaire, E. R. Unanue, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17463 (2009).
7. M. Kovacsics-Bankowski, K. L. Rock, *Science* **267**, 243 (1995).
8. A. L. Ackerman, A. Giodini, P. Cresswell, *Immunity* **25**, 607 (2006).
9. A. Schäfer, D. H. Wolf, *EMBO J.* **28**, 2874 (2009).
10. M. L. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 3029 (2008).
11. C. C. Norbury, B. J. Chambers, A. R. Prescott, H. G. Ljunggren, C. Watts, *Eur. J. Immunol.* **27**, 280 (1997).
12. A. Giodini, P. Cresswell, *EMBO J.* **27**, 201 (2008).
13. B. Arunachalam, U. T. Phan, H. J. Geuze, P. Cresswell, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 745 (2000).
14. U. T. Phan, B. Arunachalam, P. Cresswell, *J. Biol. Chem.* **275**, 25907 (2000).
15. T. Hanke, F. L. Graham, K. L. Rosenthal, D. C. Johnson, *J. Virol.* **65**, 1177 (1991).
16. See supporting material on Science Online.
17. E. E. Heldwein *et al.*, *Science* **313**, 217 (2006).
18. R. Singh, A. Jamieson, P. Cresswell, *Nature* **455**, 1244 (2008).
19. D. S. Collins, E. R. Unanue, C. V. Harding, *J. Immunol.* **147**, 4054 (1991).
20. K. L. McCoy *et al.*, *J. Immunol.* **143**, 29 (1989).
21. A. Savina *et al.*, *Cell* **126**, 205 (2006).
22. W. Zhong, P. A. Reche, C. C. Lai, B. Reinhold, E. L. Reinherz, *J. Biol. Chem.* **278**, 45135 (2003).
23. J. N. Varghese, W. G. Laver, P. M. Colman, *Nature* **303**, 35 (1983).
24. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **289**, 366 (1981).
25. Q. Ye, R. M. Krug, Y. J. Tao, *Nature* **444**, 1078 (2006).
26. N. K. Rajasagi *et al.*, *J. Virol.* **83**, 5256 (2009).
27. T. Serwold, F. Gonzalez, J. Kim, R. Jacob, N. Shastri, *Nature* **419**, 480 (2002).
28. I. Jutras, M. Desjardins, *Annu. Rev. Cell Dev. Biol.* **21**, 511 (2005).
29. N. Touret *et al.*, *Cell* **123**, 157 (2005).
30. M. Maric *et al.*, *Science* **294**, 1361 (2001).
31. This work was supported by the Howard Hughes Medical Institute and NIH grant R37AI23081 (P.C.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1394/DC1
Materials and Methods
Figs. S1 and S2
References

5 March 2010; accepted 4 May 2010
10.1126/science.1189176

Endosomal Chloride-Proton Exchange Rather Than Chloride Conductance Is Crucial for Renal Endocytosis

Gaia Novarino, Stefanie Weinert, Gesa Rickheit,* Thomas J. Jentsch†

Loss of the endosomal anion transport protein ClC-5 impairs renal endocytosis and underlies human Dent's disease. ClC-5 is thought to promote endocytosis by facilitating endosomal acidification through the neutralization of proton pump currents. However, ClC-5 is a 2 chloride (Cl[−])/proton (H⁺) exchanger rather than a Cl[−] channel. We generated mice that carry the uncoupling E211A (unc) mutation that converts ClC-5 into a pure Cl[−] conductor. Adenosine triphosphate (ATP)-dependent acidification of renal endosomes was reduced in mice in which ClC-5 was knocked out, but normal in *Clcn5*^{unc} mice. However, their proximal tubular endocytosis was also impaired. Thus, endosomal chloride concentration, which is raised by ClC-5 in exchange for protons accumulated by the H⁺-ATPase, may play a role in endocytosis.

Luminal acidification along the endocytic pathway serves many purposes (1), including the progression of endocytosis itself

(2). It is performed by endosomal H⁺-transporting adenosine triphosphatases (H⁺-ATPases) that need a countercurrent for electroneutrality. Because this

current depends on chloride, conventional wisdom suggests (1) that endosomal Cl^- channels are involved (fig. S1A). CIC-5 was thought to embody this channel in proximal tubular endosomes (3–5). Disruption of CIC-5 impairs renal endosomal acidification in vitro (5) and drastically reduces proximal tubular endocytosis in mice and humans (4, 6, 7). The hyperphosphaturia and hypercalciuria that lead to kidney stones in Dent's disease have been attributed to impaired tubular endocytosis of calcitropic hormones (4). However, it has recently been shown (8–10) that CIC-5 is a $2\text{Cl}^-/\text{H}^+$ exchanger rather than a Cl^- channel. It seems counterintuitive that such an exchanger should neutralize pump currents because it mediates H^+ -efflux during ATP-driven acidification. The biological consequence of proton coupling has remained enigmatic (11).

If the CIC-5 Cl^-/H^+ exchanger could be converted into an uncoupled Cl^- conductor, it should efficiently facilitate endosomal acidification. Phenotypes of mice carrying such a mutation cannot be attributed to impaired endosomal acidification, but can be ascribed specifically to a loss of coupling chloride gradients to proton gradients. A mutation in the “gating” glutamate of CLC exchangers (12) suffices to convert them into pure anion conductors (8, 9, 13–15). We inserted the corresponding, well-characterized E211A mutation (8, 9, 14, 16, 17) into the *Cln5* gene on mouse chromosome X and created mice in which CIC-5 was converted into an uncoupled Cl^- conductor (*Cln5^{unc}* mice) (figs. S1A and S2). The mutant protein was expressed at wild-type (WT) levels (Fig. 1A). No change was observed in its subcellular localization in kidney proximal tubular and intercalated cells (Fig. 1B and fig. S3). The renal expression of the related CIC-3 and CIC-4 proteins also was not affected (fig. S4).

To test whether the uncoupled CIC-5^{unc} mutant supported endosomal acidification, we added ATP to endosomal fractions from renal cortex (containing mainly proximal tubules) of WT or *Cln5^{unc/y}* mice and monitored vesicular pH using acridine orange fluorescence. H^+ -ATPase-driven acidification of WT and *Cln5^{unc}* vesicles occurred with similar efficiency but was severely reduced with endosomes from mice in which CIC-5 was knocked out (KO mice), as expected (5, 18) (Fig. 1, C to E).

Despite maintaining active endosomal acidification, *Cln5^{unc}* mice displayed abnormalities found in CIC-5 KO (*Cln5^{-/-}*) mice and patients with Dent's disease (4, 7, 19), such as low-molecular-weight proteinuria (Fig. 2, A and B), hyperphosphaturia, and hypercalciuria (table S1). Proteinuria of *Cln5^{-/-}* mice results from impaired

proximal tubular endocytosis (4, 7), which was studied in chimeric tubules resulting from random X-chromosomal inactivation in female *Cln5^{+/-}* mice (4). In those tubules, WT and KO cells were distinguished by means of antibodies to CIC-5 (3, 4), but this approach cannot differentiate between cells expressing the WT or uncoupled CIC-5 in *Cln5^{unc/+}* tubules. Rather than epitope-tagging the E211A mutant, which might interfere with its function, we generated mice in which the C terminus of CIC-5 was converted to that of CIC-3 (fig. S1, B and C, and S5). The generation of this *Cln5** allele required only two amino acid exchanges and changed neither CIC-5 currents (fig. S5D) nor its abundance (Fig. 1A) and localization (fig. S6A). Our antibodies against the C terminus of CIC-5 recognized CIC-5 and CIC-5^{unc} but not CIC-5* (Figs. 1A and 2, C and D, and fig. S1, B and C). In vivo endocytosis experiments were performed by injecting into the bloodstream labeled endocytic cargo that can pass the glomerular filter. Endocytosis in cells expressing the *Cln5** allele or WT CIC-5 was indistinguishable (fig. S6B). However, cells expressing the *Cln5^{unc}* allele accumulated much less fluorescently labeled β -lactoglobulin, which is a marker for receptor-mediated endocytosis (Fig. 2C), or the fluid-phase marker dextran (Fig. 2D) than did neighboring cells that express the $2\text{Cl}^-/\text{H}^+$ -exchanger CIC-5*. Thus, uncoupling anion transport from protons resulted in a cell-autonomous impairment of both receptor-mediated and fluid-phase endocytosis akin to *Cln5^{-/-}* mice (4). Similarly as in KO mice (4, 20), cells expres-

sing the uncoupled E211A mutant displayed reduced levels of the endocytic receptors megalin (Fig. 3, A and C) and cubilin (Fig. 3B)—a finding ascribed to impaired recycling to the plasma membrane (4, 20). In *Cln5^{unc}* kidney, the sodium-phosphate cotransporter NaPi-2a was shifted from the apical membrane to intracellular vesicles, and its overall abundance was reduced (fig. S7), which explains the observed hyperphosphaturia. Similarly increased endocytosis of NaPi-2a in *Cln5^{-/-}* mice was attributed to reduced endocytosis of filtered parathyroid hormone (PTH), entailing a luminal increase of PTH and excessive stimulation of apical PTH receptors (4, 5).

Because the E211A mutation did not abolish CIC-5 currents and maintained endosomal acidification (Fig. 1C), it might have affected endocytosis less than a loss of CIC-5. We thus compared *Cln5^{unc}* and *Cln5^{-/-}* cells side by side in chimeric tubules of *Cln5^{unc/+}* females. No differences were detected in receptor-mediated (Fig. 2E) or fluid-phase endocytosis (fig. S8), nor in the localization and abundance of megalin (Fig. 3C). Thus unexpectedly, the anion conductance of CIC-5(E211A) could not even partially substitute for Cl^-/H^+ exchange in the support of proximal tubular endocytosis and normal localization of apical receptors.

Because the intramembrane E211A point mutation was unlikely to have changed CIC-5 protein interactions, the pathology of *Cln5^{unc}* mice is caused neither by the disruption of a macromolecular CIC-5-containing endocytic complex (21, 22) nor by the loss of interaction with KIF3B

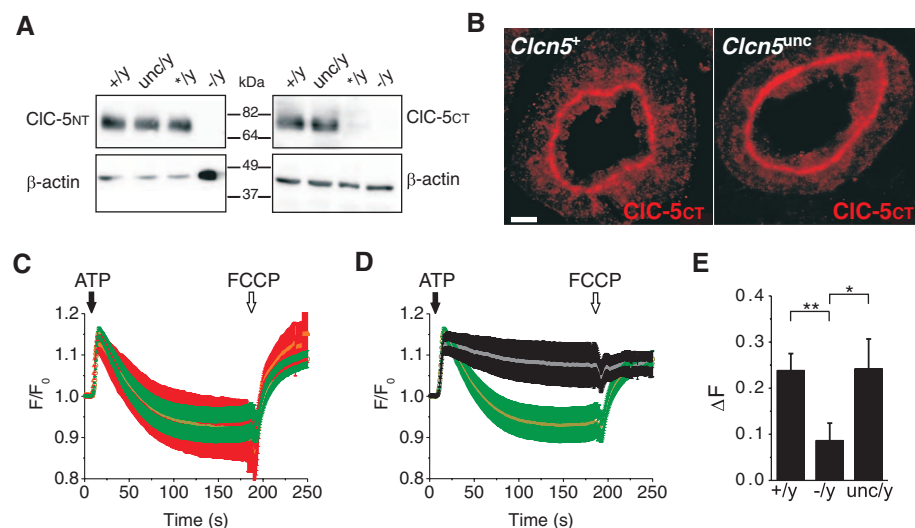


Fig. 1. Renal endosomal acidification of mice converting CIC-5 into a chloride conductor. **(A)** Immunoblot for CIC-5 using kidney membranes of different genotypes with antibodies against the (left) N terminus or the (right) C terminus (C5/O5A). Asterisk indicates mutated C terminus of CIC-5 (fig. S5). **(B)** Identical staining pattern using the C-terminal PEP5E antibody to CIC-5 (3) in proximal tubules of WT and *Cln5^{unc/y}* mice. Scale bar, 5 μm . **(C and D)** Averaged traces of acridine orange fluorescence comparing ATP-driven acidification of renal cortical vesicles from (C) WT (green) and *Cln5^{unc}* (red) and (D) WT (green) and *Cln5^{-/-}* (black) mice (with identical WT traces). Reduced fluorescence reflects vesicular acidification. The protonophore FCCP was added as a control. F_0 , fluorescence at time (t) = 0. **(E)** Averaged total change in fluorescence. Data averaged from 25 (+/y), 13 (-/-), and 10 (unc/y) experiments with material obtained from more than four independent vesicle preparations per genotype. Error bars, SEM. * $P \leq 0.05$; ** $P \leq 0.01$.

Leibniz-Institut für Molekulare Pharmakologie (FMP) and Max-Delbrück-Centrum für Molekulare Medizin (MDC), 13125 Berlin, Germany.

*Present address: TaconicArtemis GmbH, 51063 Köln, Germany.

†To whom correspondence should be addressed. E-mail: jentsch@fmp-berlin.de

(23), which are mechanisms that have been suggested to underlie Dent's disease. The impairment of endocytosis in Dent's disease is not likely to result only from reduced ATP-dependent endosomal acidification (3) but is related to an uncou-

pling of Cl^- gradients from H^+ gradients. *ClC-5* might drive H^+ -ATPase-independent acidification by exchanging cytosolic H^+ for intravesicular Cl^- at an early step of endocytosis (9). Shortly after pinching off from the plasma membrane,

endocytic vesicles may contain the high extracellular chloride concentration. A selective lack of such an initial Cl^- gradient-driven endosomal acidification might explain the impaired endocytosis of both *Clcn5^{-/-}* and *Clcn5^{unc}* mice. However, endosomal $[\text{Cl}^-]$ was found to be initially low—a finding ascribed to surface charge effects (18). Furthermore, a lack of a Cl^- gradient-driven acidification cannot be invoked for the severe lysosomal pathology of *Clcn7^{-/-}* (24, 25) and *Clcn7^{unc/unc}* mice (26) because their lysosomal pH is normal (24–26) and because this mechanism would require a substantial Cl^- supply to lysosomes.

The surprisingly similar pathologies of *Clcn5^{-/-}* and *Clcn5^{unc}* mice might be due to reduced endosomal Cl^- accumulation in each mouse model. Whereas Cl^- channels operating in parallel to an H^+ -ATPase would raise intravesicular Cl^- during active acidification, stoichiometric $2\text{Cl}^-/\text{H}^+$ exchange would maintain high vesicular Cl^- concentration also under steady state. Indeed, lysosomes containing the WT *ClC-7* Cl^-/H^+ exchanger display higher $[\text{Cl}^-]$ than their *Clcn7^{-/-}* or *Clcn7^{unc/unc}* counterparts (26). Furthermore, endosomal $[\text{Cl}^-]$ might regulate endosomal Ca^{++} channels (27). The analysis of *Clcn5^{unc}* and *Clcn7^{unc}* (26) mice suggests that luminal anion concentration is important all along the endosomal-lysosomal pathway.

Fig. 2. Proteinuria and impaired endocytosis in *Clcn5^{-/-}* *Clcn5^{unc}* mice. (A) Urinary proteins analyzed by means of SDS-polyacrylamide gel electrophoresis and silver staining. (B) Immunoblot for vitamin D-binding protein (DBP) and retinol-binding protein (RBP) in urine. (C) In vivo 10-min uptake of Alexa Fluor 546-labeled (Invitrogen, Carlsbad, CA) β -lactoglobulin (red) showed decreased receptor-mediated endocytosis in *Clcn5^{unc}*-expressing cells in a chimeric proximal tubule from a *Clcn5^{unc/+}* female. *ClC-5^{unc}*, but not the *ClC-5⁺* Cl^-/H^+ exchanger, is recognized by the antibody C5/05A (green). (Right) Overlay with additional brush-border staining for villin (blue). (D) Reduced fluid-phase endocytosis of Alexa Fluor 488-labeled (Invitrogen) dextran in *ClC-5^{unc}*-expressing cells in an experiment similar to (C). (E) Similar β -lactoglobulin-uptake of cells lacking *ClC-5* or expressing *Clcn5^{unc}* in a *Clcn5^{unc/-}* female. Scale bars, 5 μm (C), 3.2 μm (D), and 6 μm (E).

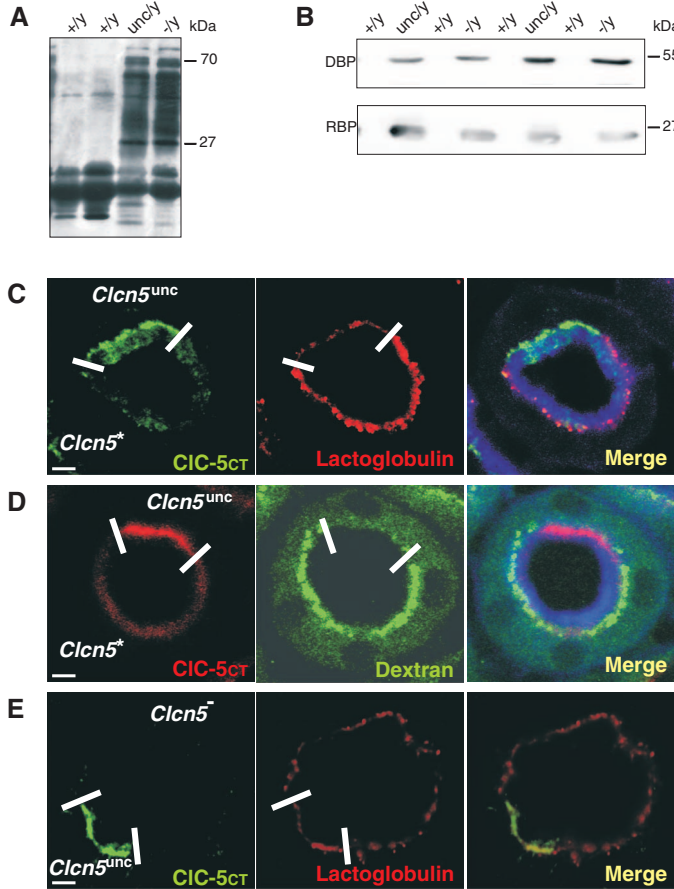
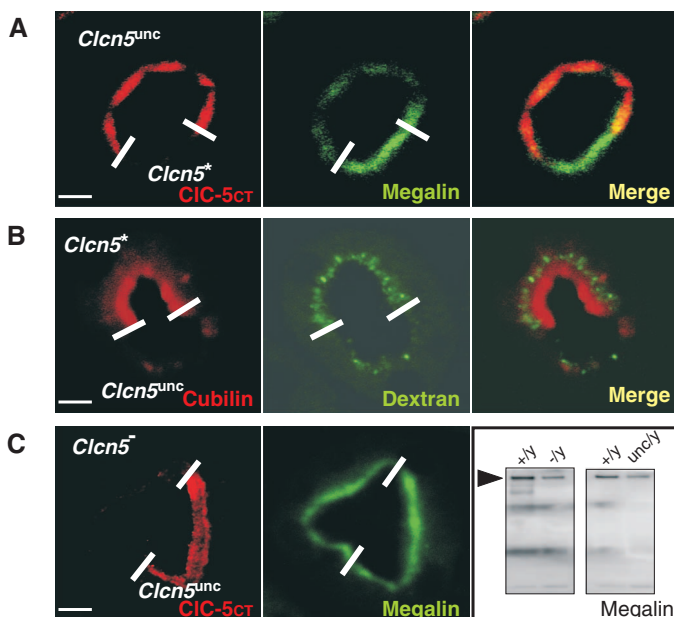


Fig. 3. Effect of uncoupling *ClC-5* on endocytic receptors. (A) Reduced megalin levels in *Clcn5^{unc}*-expressing cells of a chimeric tubule of *Clcn5^{unc/+}* mice. (B) Reduced cubilin expression in *Clcn5^{unc}*-expressing cells that were identified through reduced endocytosis of Alexa 488-dextran. (C) Similar reduction of megalin in cells expressing *Clcn5^{-/-}* or *Clcn5^{unc}*. To compare weak megalin immunostaining in *Clcn5^{unc/-}* tubules, image intensity was boosted in the center panel as compared with that of (A). (Right) Immunoblot for megalin using kidney membranes from WT, *Clcn5^{unc/+}*, and *Clcn5^{-/-}* mice. Scale bars, 4 μm [(A) and (B)] and 5 μm (C).



References and Notes

1. I. Mellman, R. Fuchs, A. Helenius, *Annu. Rev. Biochem.* **55**, 663 (1986).
2. M. J. Clague, S. Urbé, F. Aniento, J. Gruenberg, *J. Biol. Chem.* **269**, 21 (1994).
3. W. Günther, A. Lühchow, F. Cluzeaud, A. Vandewalle, T. J. Jentsch, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8075 (1998).
4. N. Piwon, W. Günther, M. Schwake, M. R. Bösl, T. J. Jentsch, *Nature* **408**, 369 (2000).
5. W. Günther, N. Piwon, T. J. Jentsch, *Pflügers Arch.* **445**, 456 (2003).
6. S. E. Lloyd *et al.*, *Nature* **379**, 445 (1996).
7. S. S. Wang *et al.*, *Hum. Mol. Genet.* **9**, 2937 (2000).
8. A. Picollo, M. Pusch, *Nature* **436**, 420 (2005).
9. O. Scheel, A. A. Zdebik, S. Lourdel, T. J. Jentsch, *Nature* **436**, 424 (2005).
10. G. Zifarelli, M. Pusch, *EMBO J.* **28**, 175 (2009).
11. T. J. Jentsch, *J. Physiol.* **578**, 633 (2007).
12. R. Dutzler, E. B. Campbell, R. MacKinnon, *Science* **300**, 108 (2003).
13. A. Accardi, C. Miller, *Nature* **427**, 803 (2004).
14. A. A. Zdebik *et al.*, *J. Biol. Chem.* **283**, 4219 (2007).
15. E. Y. Bergsdorf, A. A. Zdebik, T. J. Jentsch, *J. Biol. Chem.* **284**, 11184 (2009).
16. In the mutants, other amino acids were substituted at certain locations; for example, E211A indicates that glutamic acid at position 211 was replaced by alanine. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
17. Materials and methods are available as supporting material on Science Online.
18. M. Hara-Chikuma, Y. Wang, S. E. Guggino, W. B. Guggino, A. S. Verkman, *Biochem. Biophys. Res. Commun.* **329**, 941 (2005).
19. S. J. Scheinman, *Kidney Int.* **53**, 3 (1998).
20. E. I. Christensen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8472 (2003).
21. D. H. Hryciw *et al.*, *J. Biol. Chem.* **278**, 40169 (2003).
22. D. H. Hryciw *et al.*, *J. Biol. Chem.* **281**, 16068 (2006).
23. A. A. Reed *et al.*, *Am. J. Physiol. Renal Physiol.* **298**, F365 (2010).

24. D. Kasper *et al.*, *EMBO J.* **24**, 1079 (2005).
 25. P. F. Lange, L. Wartosch, T. J. Jentsch, J. C. Fuhrmann, *Nature* **440**, 220 (2006).
 26. S. Weinert *et al.*, *Science* **328**, 1401 (2010); published online 29 April 2010 (10.1126/science.1188072).
 27. M. Saito, P. I. Hanson, P. Schlesinger, *J. Biol. Chem.* **282**, 27327 (2007).
 28. We thank R. Pareja-Alcaraz and S. Rode for technical assistance and P. Verroust, S. Bachmann,

W. Blaner, and C. Wagner for antibodies against cubilin, megalin, retinol binding protein, and AE1, respectively. This work was supported by the Deutsche Forschungsgemeinschaft (grants Zd58/1, Je164/6, and Je164/9).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1188070/DC1
 Materials and Methods

Figs. S1 to S8
 Table S1
 References

8 February 2010; accepted 21 April 2010
 Published online 29 April 2010;
 10.1126/science.1188070
 Include this information when citing this paper.

Lysosomal Pathology and Osteopetrosis upon Loss of H⁺-Driven Lysosomal Cl⁻ Accumulation

Stefanie Weinert,^{1,2} Sabrina Jabs,^{1,2,6} Chayarop Supanchart,³ Michaela Schweizer,⁴ Niclas Gimber,^{1,2} Martin Richter,^{1,6} Jörg Rademann,^{1,6*} Tobias Stauber,^{1,2} Uwe Kornak,^{3,5} Thomas J. Jentsch^{1,2†}

During lysosomal acidification, proton-pump currents are thought to be shunted by a chloride ion (Cl⁻) channel, tentatively identified as CIC-7. Surprisingly, recent data suggest that CIC-7 instead mediates Cl⁻/proton (H⁺) exchange. We generated mice carrying a point mutation converting CIC-7 into an uncoupled (unc) Cl⁻ conductor. Despite maintaining lysosomal conductance and normal lysosomal pH, these *Clcn7^{unc/unc}* mice showed lysosomal storage disease like mice lacking CIC-7. However, their osteopetrosis was milder, and they lacked a coat color phenotype. Thus, only some roles of CIC-7 Cl⁻/H⁺ exchange can be taken over by a Cl⁻ conductance. This conductance was even deleterious in *Clcn7^{+/-unc}* mice. *Clcn7^{+/-}* and *Clcn7^{unc/unc}* mice accumulated less Cl⁻ in lysosomes than did wild-type mice. Thus, lowered lysosomal chloride may underlie their common phenotypes.

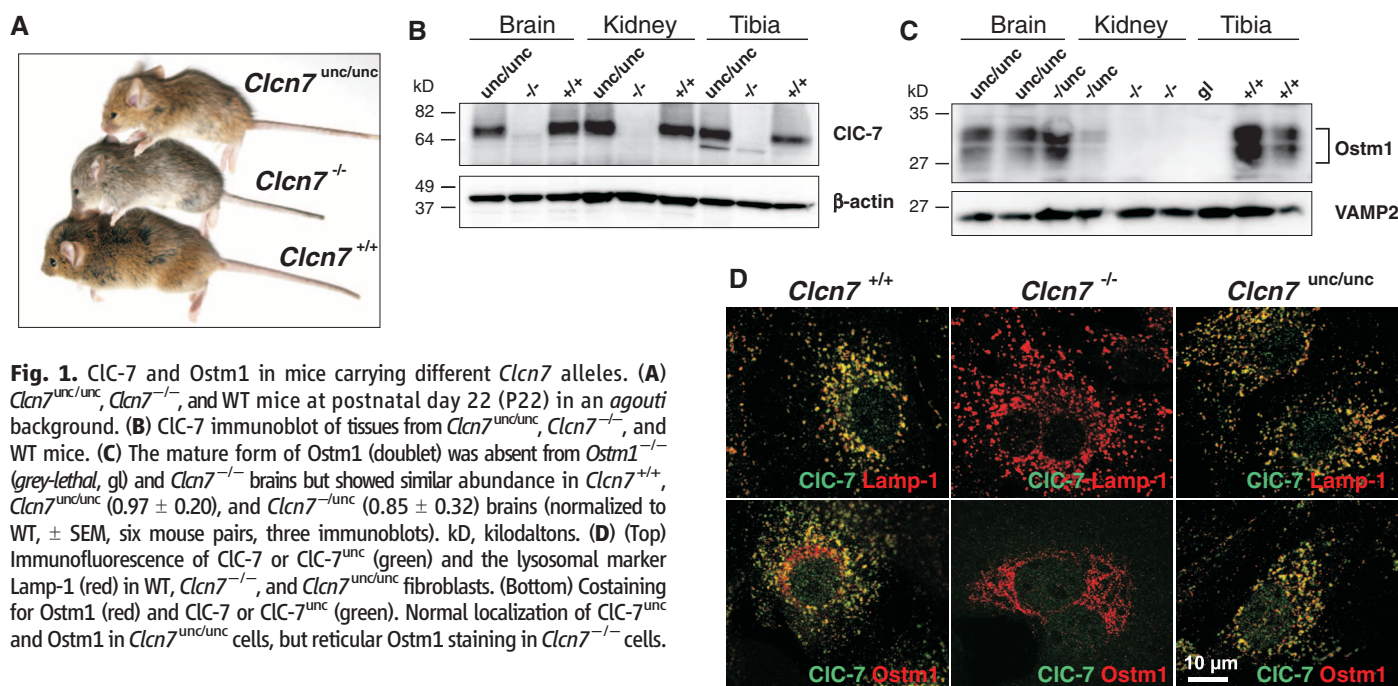
CIC-7 is the only member of the CLC gene family of anion transporters substantially expressed on lysosomes (1–3), where it resides together with its β -subunit Ostm1 (3). Inactivation of either subunit leads to lysosomal storage disease and osteopetrosis in mice and humans (1–4). Cellular defects include slowed degradation of endocytosed proteins (5) and impaired acidification of the osteoclast resorption lacuna (1). Cl⁻

currents mediated by CIC-7 have been deemed necessary for shunting lysosomal proton-pump currents (1). However, lysosomal pH was normal in cells lacking either CIC-7 or Ostm1 (2, 3). CIC-7 now seems likely to be a Cl⁻/H⁺ exchanger rather than a Cl⁻ channel (6, 7). Because H⁺-pump currents may be neutralized by both Cl⁻ channels and electrogenic Cl⁻/H⁺ exchangers (6), it is unclear whether lysosomal Cl⁻/H⁺ exchange confers functional

advantages over the simple Cl⁻ conductance in the textbook model for vesicular acidification.

We created knock-in mice in which the CIC-7 “gating” glutamate (E) was mutated to alanine (A) (fig. S1) (8). On the basis of results from other CLC Cl⁻/H⁺ exchangers (9–12), this Glu²⁴⁵ → Ala²⁴⁵ (E245A) mutation should lead to Cl⁻ transport that is uncoupled (unc) from protons, hence our designation of this allele as *Clcn7^{unc}*. Homozygous *Clcn7^{unc/unc}* mice showed severe growth retardation (Fig. 1A and fig. S2) and died within 5 weeks. *Clcn7^{unc}* and wild-type (WT) CIC-7 were expressed at similar levels (Fig. 1B) and similarly localized to lysosomes (Fig. 1D). Neither the abundance, nor the lysosomal localization of Ostm1 was changed in *Clcn7^{unc/unc}* mice, contrasting with its strongly reduced protein level (3) and mislocalization in *Clcn7^{+/-}* cells (Fig. 1, C and D). In neurons, however, *CIC-7^{unc}* staining was more diffuse (fig. S3B), reflecting changed lysosomal compartments like in *Clcn7^{+/-}* neurons (2). The abundance of other CLC exchangers was unchanged in *Clcn7^{unc/unc}* mice (fig. S4).

In an *agouti* genetic background, the coat color of *Clcn7^{+/-}* and *Ostm1^{+/-}* (*grey-lethal*) mice is grey (3, 4), whereas it was brownish in WT and *Clcn7^{unc/unc}* mice (Fig. 1A). *Clcn7^{unc/unc}* mice were osteopetrotic (Fig. 2A and fig. S5), although less severely than *Clcn7^{+/-}* (1) or *Ostm1^{+/-}* (4) mice. CIC-7 and Ostm1 were detected at the ruffled border of *Clcn7^{unc/unc}* osteoclasts (fig. S3A). This



Covering a Broad Dynamic Range: Information Processing at the Erythropoietin Receptor

Verena Becker,^{1,2*} Marcel Schilling,^{1*} Julie Bachmann,¹ Ute Baumann,¹ Andreas Raue,³ Thomas Maiwald,^{3†} Jens Timmer,^{3,4,5} Ursula Klingmüller^{1,2‡}

Cell surface receptors convert extracellular cues into receptor activation, thereby triggering intracellular signaling networks and controlling cellular decisions. A major unresolved issue is the identification of receptor properties that critically determine processing of ligand-encoded information. We show by mathematical modeling of quantitative data and experimental validation that rapid ligand depletion and replenishment of the cell surface receptor are characteristic features of the erythropoietin (Epo) receptor (EpoR). The amount of Epo-EpoR complexes and EpoR activation integrated over time corresponds linearly to ligand input; this process is carried out over a broad range of ligand concentrations. This relation depends solely on EpoR turnover independent of ligand binding, which suggests an essential role of large intracellular receptor pools. These receptor properties enable the system to cope with basal and acute demand in the hematopoietic system.

Cells respond to alterations in their environment that are frequently encoded by changes in the concentration of extracellular ligands. These changes are perceived by cell surface receptors, and the high frequency with which receptors are mutated in diseases and their accessibility to drugs make them key targets for therapeutic interventions. The dynamics of receptor activation are critically determined by the capacity to capture and sequester ligand through endocytosis (1). Ligand-encoded information could be processed in a saturation-like or linear mode for increasing ligand concentrations (Fig. 1A). Receptor proper-

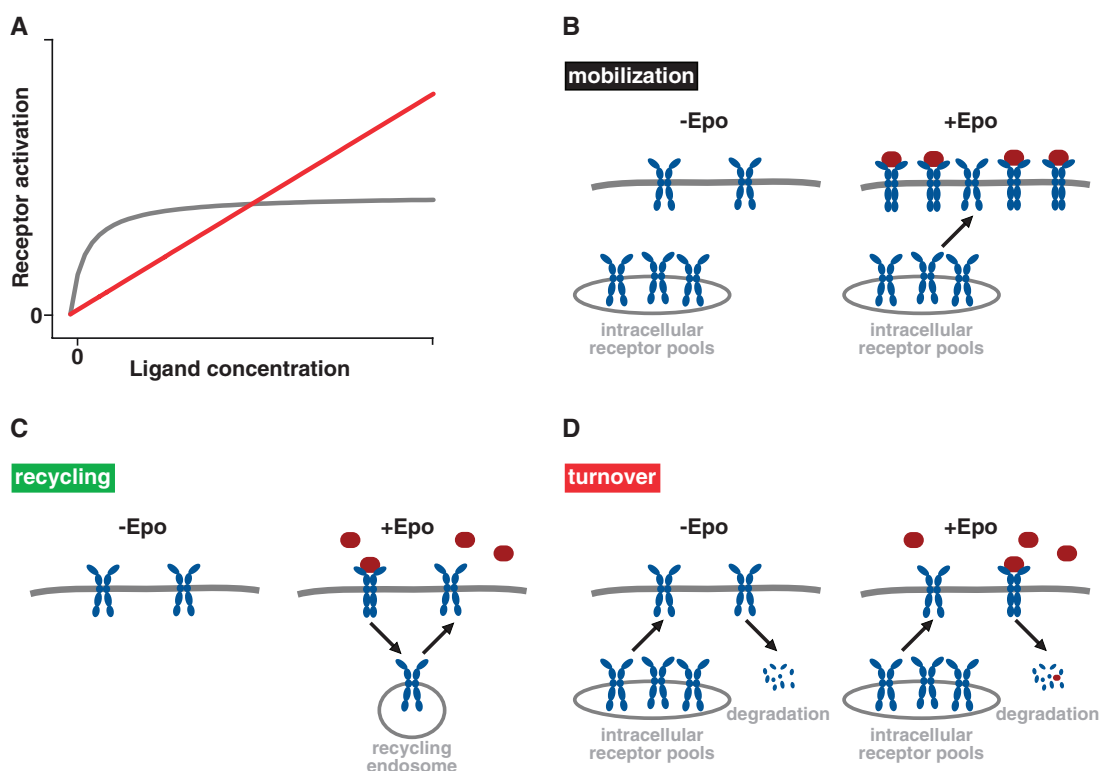
ties enabling cells to cope with ligand concentrations that vary over a broad range remained to be identified.

A prime example for a receptor that encounters an extreme range of ligand concentrations is the erythropoietin receptor (EpoR). The EpoR ensures continuous renewal of short-lived erythrocytes (2) and enhanced expansion of erythroid progenitors upon demands such as blood loss. Plasma concentrations of erythropoietin (Epo) can differ ~1000-fold between basal and acute conditions (3). Only a small proportion of EpoR is present on the cell surface; the majority resides in intracellular pools (4). Ligand binding triggers

phosphorylation of the cytoplasmic EpoR domain by Janus kinase 2 (JAK2) (5). Ligand-induced endocytosis of EpoR has been proposed to terminate signaling by removing receptors from the cell surface (6). Additionally, the EpoR is subjected to ligand-independent endocytosis (7) [supporting online material (SOM) text]. The specific impact of EpoR transport to the plasma membrane and EpoR endocytosis on processing of ligand-encoded information have remained ill-defined.

Linear detection of ligand over a broad range of concentrations could be facilitated by the following properties reported for other receptors (Fig. 1, B to D, and SOM text): (i) mobilization, defined as ligand-induced additional transport of newly synthesized receptor from intracellular pools to the plasma membrane (8); (ii) recycling, consisting of ligand-induced receptor endocytosis and subsequent transport back to the plasma membrane (9); and (iii) turnover, comprising ligand-independent transport of newly synthesized receptor to the plasma membrane and removal from the plasma membrane by ligand-

Fig. 1. Strategies of information processing through cell surface receptors for a broad range of ligand concentrations. **(A)** Representation of two modes for information processing. **(B to D)** Hypothetical mechanisms for linear information processing through the EpoR: mobilization, recycling, and turnover (see text).



¹Division Systems Biology of Signal Transduction, DKFZ-ZMBH Alliance, German Cancer Research Center, 69120 Heidelberg, Germany. ²Bioquant, Heidelberg University, 69120 Heidelberg, Germany. ³Institute of Physics, University of Freiburg, 79104 Freiburg, Germany. ⁴Freiburg Institute for Advanced Studies, University of Freiburg, 79104 Freiburg, Germany. ⁵Centre for Biological Signalling Studies (BIOS), University of Freiburg, 79104 Freiburg, Germany.

*These authors contributed equally to this work.

†Present address: Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

‡To whom correspondence should be addressed. E-mail: u.klingmueller@dkfz-heidelberg.de

independent receptor endocytosis and subsequent degradation (10). A further potential strategy to cope with particularly high ligand concentrations is expression of large amounts of receptor on the plasma membrane (11). However, the abundance of EpoR on the cell surface is rather low (12) (fig. S1 and SOM text).

Receptor mobilization, recycling, and turnover are highly dynamic and intertwined processes that are difficult to disentangle experimentally. To address these nonlinear processes, we developed dynamic mathematical models for ligand-receptor interaction and trafficking kinetics and fitted them to quantitative experimental data (13).

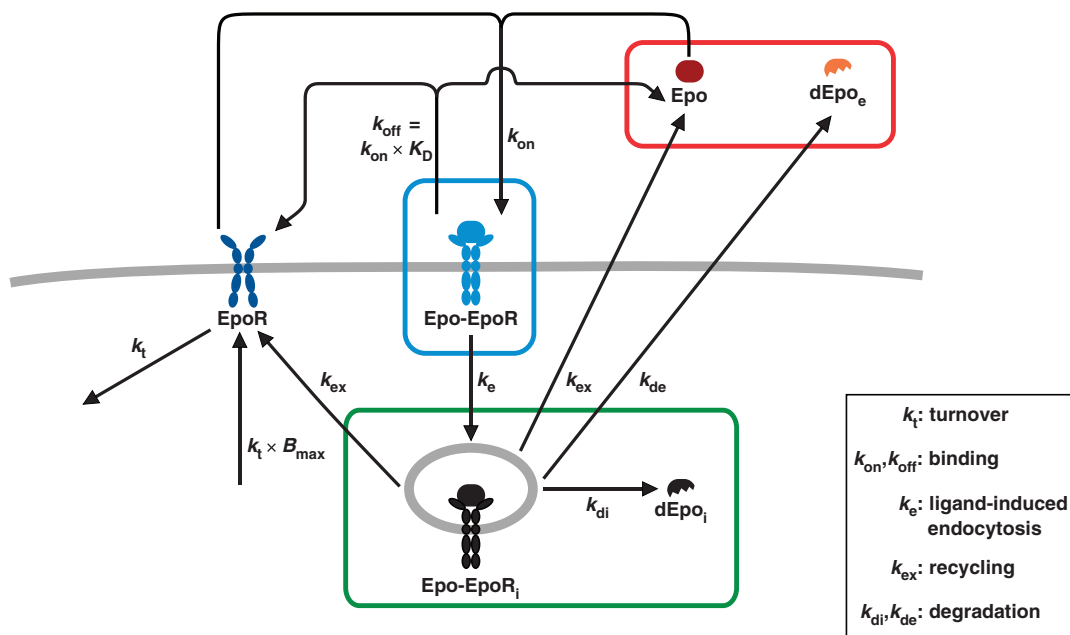
The core model included parameters for EpoR recycling and turnover (Fig. 2A) and was compared with an extended model encompassing receptor mobilization (core model + k_{mob}) (fig. S5A). Briefly, unoccupied cell surface EpoR is subjected to turnover with transport of newly synthesized receptor to the plasma membrane ($k_i \times B_{\text{max}}$)—where B_{max} is the maximal binding capacity of the total cell surface receptor population—and ligand-independent endocytosis (k_t). Epo binds to cell surface receptor with the association rate k_{on} and dissociates with the rate k_{off} . The definition of the parameter k_{off} is based on the dissociation constant K_D ($k_{\text{on}} \times K_D$). Epo-EpoR complexes

are subjected to endocytosis (k_e). These complexes dissociate, and EpoR and Epo recycle back to the plasma membrane (k_{ex}) or undergo degradation. Degraded Epo is retained in intracellular compartments (k_{di}) or released to the extracellular space in an inactive state (k_{de}), unable to rebind to EpoR. In our extended model, we additionally integrated EpoR mobilization as a single parameter k_{mob} to summarize its overall effect, including a chaperone action mediated by the protein kinase JAK2 (14) (fig. S2A).

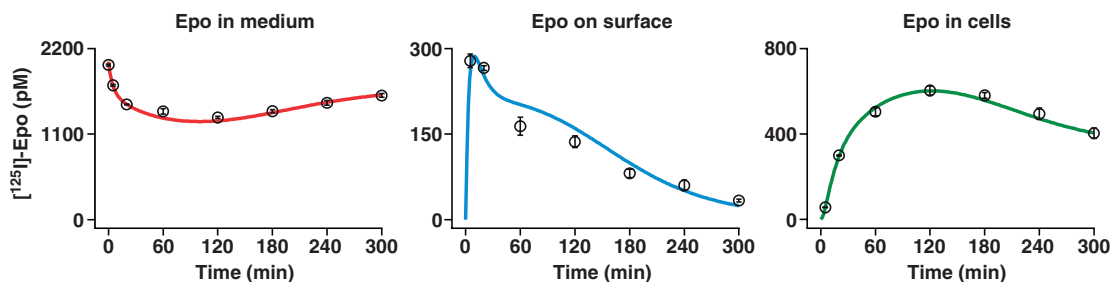
We calibrated the mathematical model on the basis of experimental data from BaF3-EpoR cells, a murine proB cell line that exogenously

Fig. 2. Dynamic modeling of the EpoR system. (A) Graphical representation of the mathematical core model. Colored boxes indicate the experimentally accessible quantities “Epo in medium” (Epo + dEpo_e, red), “Epo on surface” (Epo-EpoR, blue), and “Epo in cells” (Epo-EpoR_i + dEpo_i, green). Biological processes described by individual reaction rates are given in the inset. (B) Global parameter estimation was performed simultaneously for the core model and the auxiliary model (fig. S4). Experimental data for the core model are represented with standard deviations ($n = 3$), and trajectories of the best fit are shown. (C) Trajectories for the predicted behavior of experimentally unobserved dynamic variables of the core model.

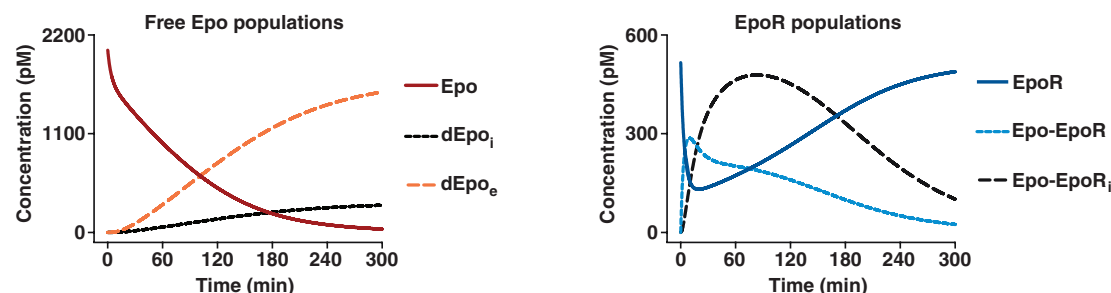
A Mathematical core model



B Model calibration



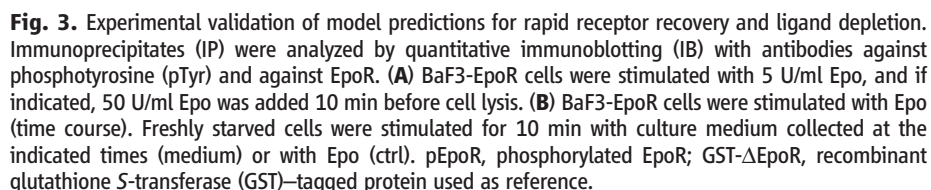
C Dynamic model predictions



To confirm experimentally the predicted recovery of cell surface EpoR, we performed time-course analysis of receptor activation in BaF3-EpoR cells. Receptor phosphorylation returned to basal amounts between 60 and 120 min after cells were exposed to Epo (Fig. 3A). However, restimulation of these cells with an

Model simulations for increasing ligand concentrations showed saturation for the peak amplitude of cell surface Epo-EpoR complexes (Fig. 4A, middle) that resulted from the limited amount of receptor present on the plasma membrane at a given time. A linear relation for integral EpoR occupancy representing the amount of cell surface Epo-EpoR complexes integrated over time was

In this study, we identified rapid ligand depletion and compensation of endocytic removal of cell surface EpoR by receptor turnover as hallmarks of the EpoR that facilitate linear information processing for a broad range of ligand concentrations. These systems properties enable cells to sample extracellular ligand continuously and there-

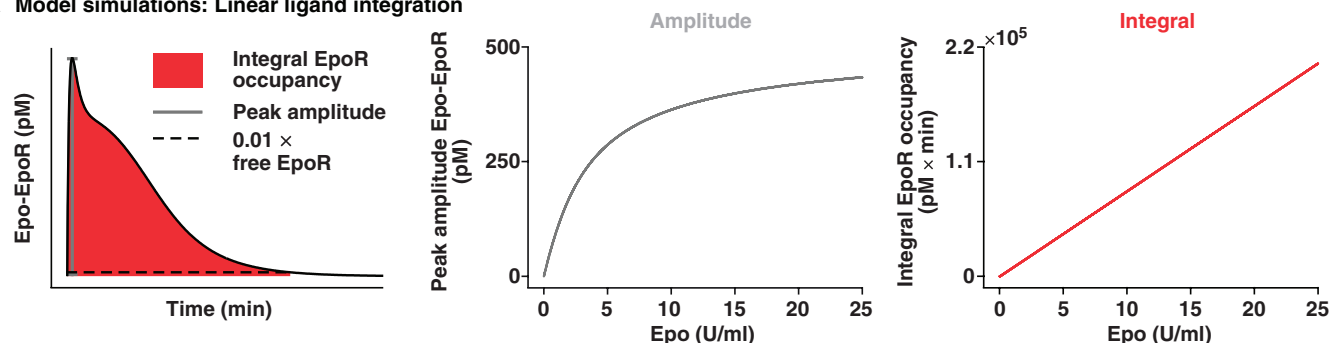


by to cope with both basal and acute demand-driven ligand concentrations. Epo has been widely applied to treat anemia (19), and current research focuses on engineering more efficient erythropoiesis-

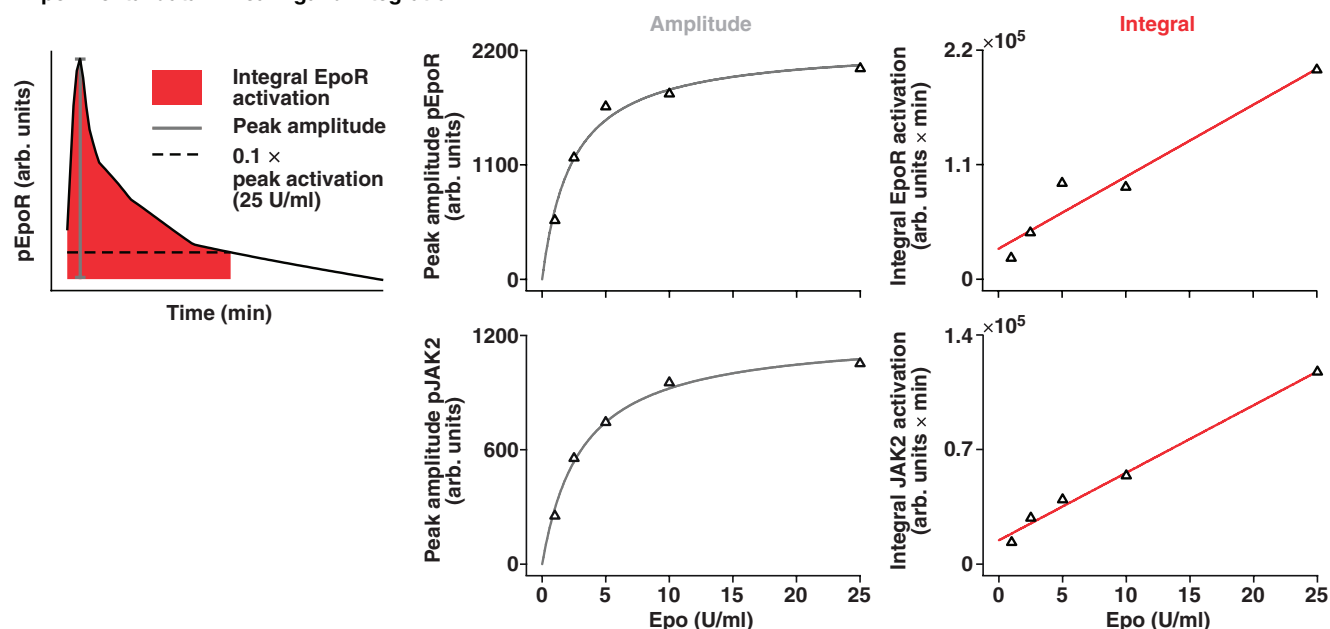
stimulating agents (20). Model simulations for various ligand-binding rates point to a parameter region that displays a trade-off between bioavailability and bioactivity of Epo derivatives (figs.

S14 to S16 and SOM text). Thus, the combination of mathematical modeling and quantitative biochemical analysis enables a more rational development of therapeutic agents.

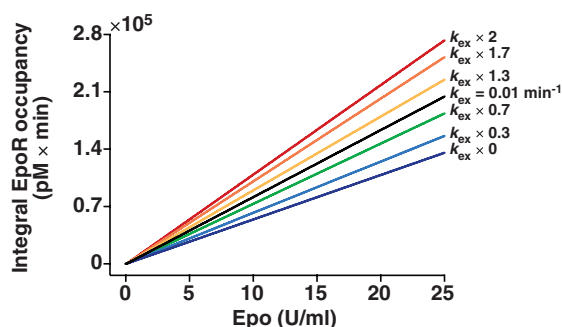
A Model simulations: Linear ligand integration



B Experimental data: Linear ligand integration



C Model simulations: Dependence on k_{ex}



D Model simulations: Dependence on k_t

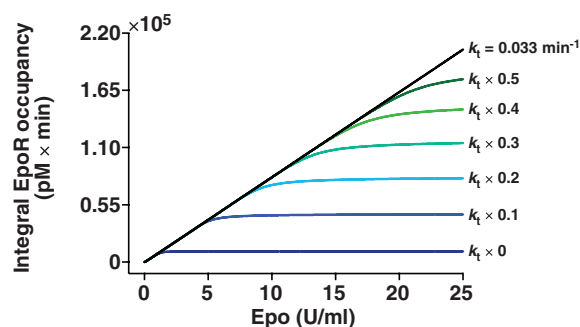


Fig. 4. Linear conversion of Epo concentrations into integral receptor occupancy depends on EpoR turnover. (A) Simulations for peak amplitude of Epo-EpoR complexes at the plasma membrane (middle) and integral EpoR occupancy defined as the amount of cell surface Epo-EpoR complexes integrated over time relative to their dependence on Epo (right). Graphical representations are shown for peak amplitude and integral EpoR occupancy (left). (B) Experimental data (triangles) were acquired by quantitative immunoblot analysis of EpoR and JAK2 activation in BaF3-EpoR cells (figs.

S10 and S11). A Michaelis-Menten-like saturation and a linear function were fitted to the data for peak amplitude (middle) and integral EpoR and JAK2 activation, defined as the amount of phosphorylated protein integrated over time (right), respectively. Graphical representations are shown for peak amplitude and integral EpoR activation (left). (C and D) Simulations for integral EpoR occupancy for various parameter values around the estimated rates of (C) recycling k_{ex} or (D) turnover k_t . Details for calculating the integral are provided in the SOM text. pEpoR and pJAK2, phosphorylated proteins.

References and Notes

- H. Shankaran, H. Resat, H. S. Wiley, *PLOS Comput. Biol.* **3**, e101 (2007).
- H. Wu, X. Liu, R. Jaenisch, H. F. Lodish, *Cell* **83**, 59 (1995).
- W. Jelkmann, *Intern. Med.* **43**, 649 (2004).
- A. Yoshimura, A. D. D'Andrea, H. F. Lodish, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 4139 (1990).
- T. D. Richmond, M. Chohan, D. L. Barber, *Trends Cell Biol.* **15**, 146 (2005).
- P. Walrafen *et al.*, *Blood* **105**, 600 (2005).
- D. L. Beckman, L. L. Lin, M. E. Quinones, G. D. Longmore, *Blood* **94**, 2667 (1999).
- K. S. Price *et al.*, *Am. J. Respir. Cell Mol. Biol.* **28**, 420 (2003).
- Z. Bajzer, A. C. Myers, S. Vuk-Pavlović, *J. Biol. Chem.* **264**, 13623 (1989).
- S. Belouard, D. Delcroix, Y. Rouillé, *J. Biol. Chem.* **279**, 28499 (2004).
- G. Carpenter, S. Cohen, *Annu. Rev. Biochem.* **48**, 193 (1979).
- H. Youssoufian, G. Longmore, D. Neumann, A. Yoshimura, H. F. Lodish, *Blood* **81**, 2223 (1993).
- Materials and methods are available as supporting material on *Science* Online.
- L. J. Huang, S. N. Constantinescu, H. F. Lodish, *Mol. Cell* **8**, 1327 (2001).
- T. Maiwald, J. Timmer, *Bioinformatics* **24**, 2037 (2008).
- A. Raue *et al.*, *Bioinformatics* **25**, 1923 (2009).
- H. Resat, J. A. Ewald, D. A. Dixon, H. S. Wiley, *Biophys. J.* **85**, 730 (2003).
- H. S. Wiley *et al.*, *J. Biol. Chem.* **266**, 11083 (1991).
- W. Jelkmann, *Eur. J. Haematol.* **78**, 183 (2007).
- I. C. Macdougall, K. U. Eckardt, *Lancet* **368**, 947 (2006).
- We thank R. Eils, U. Kummer, R. Meyer, A. C. Pfeifer, J. P. Schlöder, C. Schultz, and V. Starkuviene for critically reading the manuscript; S. N. Constantinescu for HA-EpoR constructs; and S. Manthey and S. Lattermann for technical assistance. This work was supported by the Helmholtz Alliance on Systems Biology (SB Cancer) (V.B., M.S., J.T., and U.K.), the German Federal Ministry of Education and Research (BMBF)-funded MedSys-Network LungSys (J.B., A.R., and U.K.), and the Excellence Initiative of the German Federal and State Governments (EXC 294) (J.T.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1184913/DC1
Materials and Methods

SOM Text

Figs. S1 to S16

References

19 November 2009; accepted 5 May 2010

Published online 20 May 2010;

10.1126/science.1184913

Include this information when citing this paper.

The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans

Carsten K. W. De Dreu,^{1*} Lindred L. Greer,¹ Michel J. J. Handgraaf,¹ Shaul Shalvi,¹ Gerben A. Van Kleef,¹ Matthijs Baas,¹ Femke S. Ten Velden,¹ Eric Van Dijk,² Sander W. W. Feith³

Humans regulate intergroup conflict through parochial altruism; they self-sacrifice to contribute to in-group welfare and to aggress against competing out-groups. Parochial altruism has distinct survival functions, and the brain may have evolved to sustain and promote in-group cohesion and effectiveness and to ward off threatening out-groups. Here, we have linked oxytocin, a neuropeptide produced in the hypothalamus, to the regulation of intergroup conflict. In three experiments using double-blind placebo-controlled designs, male participants self-administered oxytocin or placebo and made decisions with financial consequences to themselves, their in-group, and a competing out-group. Results showed that oxytocin drives a “tend and defend” response in that it promoted in-group trust and cooperation, and defensive, but not offensive, aggression toward competing out-groups.

Intergroup conflict is among the most pervasive problems facing human society, giving rise to such phenomena as prejudice, terrorism, ethnic cleansing, and interstate war (1, 2). Results can be devastating: Governmental genocidal policies killed more than 210 million people during the 20th century alone, and since 2000 more than 30,000 people have been killed by terrorists (3). Individuals contribute to these atrocities and their less violent but no less pervasive counterparts through parochial altruism: They self-sacrifice (i) to benefit their own group (“in-group love”) and (ii) to derogate, hurt, and sabotage competing out-groups (“out-group aggression”) (4, 5). As in-group love furthers the power and effectiveness of one’s own group vis-à-vis the competing out-group, in-group love is an indirect way of competing with the out-group.

Out-group aggression undermines the out-group’s power and effectiveness and thus is an indirect form of cooperation toward one’s own group that is often honored and publicly recognized by in-group leaders as heroic, loyal, and patriotic behavior (6, 7). Out-group aggression may be driven by the desire to increase the in-group’s relative status and power in the intergroup competition (henceforth “out-group hate”). Alternatively, it may be driven by the vigilant desire to defend and protect the in-group against real or perceived out-group threat (8, 9).

Parochial altruism figures prominently in evolutionary explanations of human social behavior. As noted by Darwin (p. 156) (10), “groups with a greater number of courageous, sympathetic and faithful members, who were always ready to warn each other of danger, to aid and defend each other... would spread and be victorious over other tribes.” The pivotal implication is that the human brain evolved to sustain motivated cognition and behavior critical to the survival of one’s own group, to facilitate contributions to in-group welfare, and to defend against outside threats, including competing groups (1). We examine whether

parochial altruism has its biological basis in brain oxytocin—a peptide of nine amino acids that is produced in the hypothalamus and released into both the brain and the bloodstream (11). Functioning as both a neurotransmitter and a hormone, oxytocin’s targets are widespread and include the amygdala, hippocampus, brainstem, and regions of the spinal cord that regulate the autonomic nervous system (12). Its manifold effects include the promotion of trust and cooperation. For example, affiliating with close kin associates with the release of blood plasma oxytocin (13), and larger numbers of oxytocin receptors (OXTR) in the human brain associate with greater empathy, generosity, and other-regarding preferences (12, 14). Finally, exogenous oxytocin (e.g., administered through nasal spray) promotes general trust and cooperation and reduces the tendency to take advantage of others’ cooperation (15–18).

Oxytocin has not been implicated in the way humans regulate intergroup competition and conflict, and perhaps oxytocin stimulates trust and cooperation toward other in-group and out-group members alike. However, compared with the interpersonal exchanges in which oxytocin has been studied thus far, cooperation takes on a radically different purpose and meaning in intergroup competition and conflict. As noted, cooperation directed toward the in-group functions to preserve, defend, and strengthen the in-group and indirectly reduces the effectiveness and power of competing out-groups; noncooperation and aggression directed at the out-group hurts the out-group and indirectly protects and strengthens the in-group (4–7). Accordingly, we hypothesized that when humans are organized in in-groups and competing out-groups, oxytocin modulates parochial altruism. It increases (i) in-group trust—the positive expectation that in-group members self-sacrifice to promote in-group welfare, (ii) in-group love, (iii) out-group hate—the inclination to aggress against the out-group to increase relative standing, and (iv) defensive out-group aggression—hostility aimed at warding off out-group threat (e.g., preemptive strike). The latter two hypotheses resonate with the notion that aggression against the out-

¹Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, Netherlands. ²Department of Psychology, Leiden University, Postbox 9555, 2300 RB, Netherlands. ³Stichting AllesKits, Cyprusaan 410, 3059 XA Rotterdam, Netherlands.

*To whom correspondence should be addressed. E-mail: c.k.w.dedreu@uva.nl

group indirectly promotes the welfare of one's own group (6, 7), that oxytocin in nonhuman mammals promotes territoriality and aggression toward intruders (19–21), and that exogenous oxytocin in humans stimulates envy when interpersonal competition is lost and gloating when the game is won (22).

To address these issues, we designed three experiments. Experiments 1 and 2 tested whether oxytocin stimulates in-group love, out-group hate, or both. Experiment 2 addressed whether the effects of oxytocin are limited to individuals with cooperative personalities, who may respond more strongly to oxytocin than individuals with noncooperative personalities (12, 14). Experiment 3 manipulated out-group threat to examine oxytocin's influence on preemptive strikes, a form of defense-motivated aggression. All experiments were computer-mediated and used a double-blind, placebo-controlled design in which participants received intranasal administration of oxytocin (or placebo containing the carrier with-

out the neuropeptide). Thirty minutes later, participants were assigned, on the basis of a trivial criterion, to two three-person groups (the group they were assigned to being the in-group, the other being the out-group) and introduced to an intergroup game in which they made confidential decisions that had financial consequences for themselves, their fellow in-group members, and the competing out-group (23).

In Experiment 1, 49 healthy males were introduced to the intergroup prisoners' dilemma-maximizing differences game (IPD-MD) (23, 24). Each individual was given €10. Each Euro kept was worth €1 for the individual; each Euro contributed to the within-group pool added €0.50 to each in-group member, including the contributor; each Euro contributed to the between-group pool added €0.50 to each in-group member, including the contributor and, in addition, it subtracted €0.50 from each out-group member. The game captures those intergroup conflicts in which contributing nothing yields the highest personal outcomes regardless of what others do. Contributing to the within-group pool yields the highest benefit to the in-group (and the larger collective) and thus reflects a cooperative motivation to benefit the in-group without hurting the out-group (in-group love). Contributing to the between-group pool, in contrast, reflects spiteful out-group hate.

Participants indicated how much of their €10 they contributed to the within-group pool, to the between-group pool, and how much they kept for themselves. Contributions to the within-group and between-group pools were submitted to a 2 (treatment: oxytocin/placebo) $\times$ 2 (pool: within/between) mixed-model analysis of variance (ANOVA). Results showed that in-group love exceeded out-group hate [$F_{1,47} = 5.54, P < 0.025$]. This effect was qualified by a treatment $\times$ pool interaction [$F_{1,47} = 4.38, P < 0.05$]: Oxytocin amplified in-group love [$F_{1,47} = 8.18, P < 0.001$] and neither increased nor decreased spiteful out-group hate [$F_{1,47} < 1$] (Fig. 1A). This also follows from an analysis in which we classified participants according to their dominant strategy. Those who gave more to

themselves than to in-group love or out-group hate were classified as egoists. In-group lovers gave more to in-group love than to out-group hate or themselves. Out-group haters gave more to the between-group pool than to in-group love or themselves. Oxytocin influenced the distribution of participants across these three strategies [Cramer's $V = 0.431$ ($N = 49$), $P < 0.011$]. In the placebo condition, 52% pursued the egoist strategy, and only 20% were in-group lovers. In the oxytocin condition, however, 17% of the participants were egoistic and 58% were in-group lovers. The number of out-group haters was similar in the oxytocin (25%) and placebo (28%) conditions. Thus, oxytocin led people away from shortsighted selfishness to in-group love but did not affect spiteful out-group hate.

After allocation decisions, participants stated whether they expected (i) their in-group members to contribute to the within-group pool (in-group trust) and (ii) out-group members to contribute to the between-group pool (out-group distrust). A 2 (treatment: oxytocin/placebo) $\times$ 2 (focus: in-group trust/out-group distrust) ANOVA showed that in-group trust exceeded out-group distrust [$F_{1,47} = 19.90, P < 0.001$]. A treatment $\times$ focus interaction [$F_{1,47} = 6.81, P < 0.015$] showed that oxytocin increased in-group trust [$F_{1,47} = 6.56, P < 0.015$] and neither increased nor decreased out-group distrust [$F_{1,47} = 2.47, P < 0.13$] (Fig. 1B).

Experiment 1 revealed that oxytocin influenced participants' allocation strategy. Because humans differ in their natural inclination to cooperate (25), perhaps cooperative individuals respond more strongly to oxytocin than noncooperative individuals (12, 14). Experiment 2 examined whether effects of oxytocin on parochial altruism found in Experiment 1 generalize across cooperative and noncooperative individuals or remain limited to cooperative individuals. Before oxytocin administration, 67 males completed the standard social value orientations test in which they chose nine times between distributions of outcomes to oneself and an anonymous other (25). An example is the choice between (i) 40 to self and 40 to other (the cooperative choice), (ii)

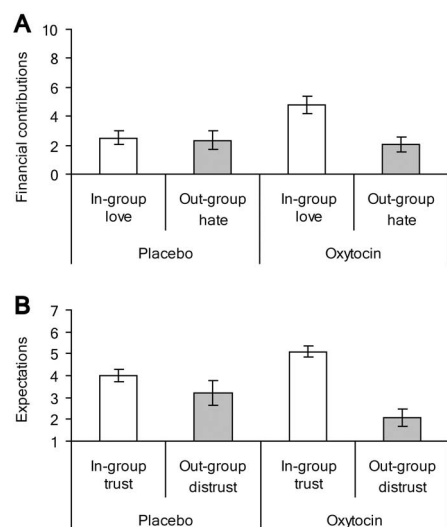


Fig. 1. (A) Financial contributions made in the IPD-MD game (displayed $\pm$ SE). **(B)** Participant's in-group trust and out-group distrust [measured on a seven-point Likert scale, ranging from 1 (low) to 7 (high), displayed $\pm$ SE].

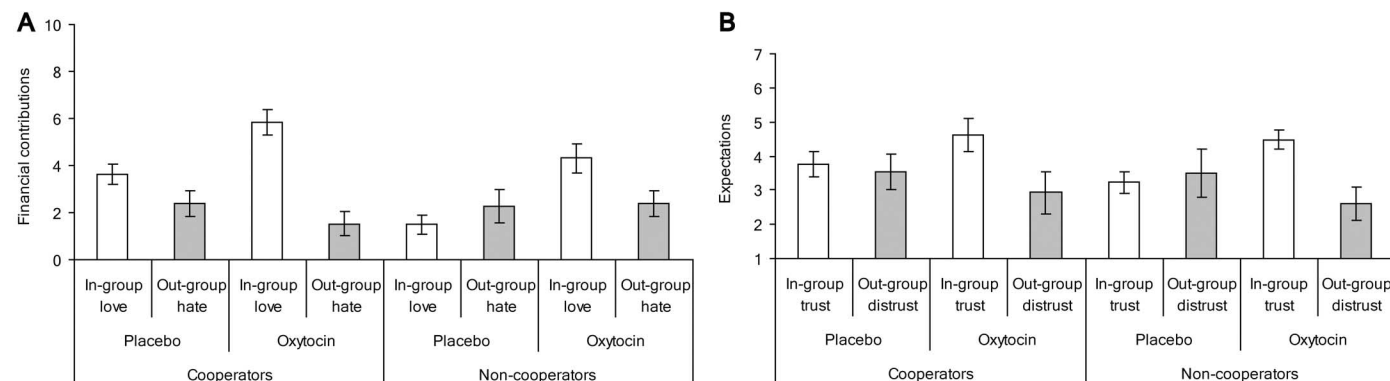


Fig. 2. (A) Financial contributions made in the IPD-MD game (displayed $\pm$ SE). **(B)** Participant's in-group trust and out-group distrust [ranging from 1 (low) to 7 (high), displayed $\pm$ SE].

50 to self and 20 to other, and (iii) 40 to self and 0 to other (ii and iii are noncooperative choices) (23, 25). Participants were classified as cooperator if they made at least six out of nine cooperative choices ($N = 25$) or as noncooperator if they made at least six out of nine noncooperative choices ($N = 42$) (23).

Experiment 2 was otherwise identical to Experiment 1. Contributions to the within-group and between-group pools were analyzed in a 2 (treatment: oxytocin/placebo) $\times$ 2 (disposition: cooperators/noncooperators) $\times$ 2 (pool: within/between) ANOVA. This revealed effects for treatment [$F_{1,63} = 6.62, P < 0.012$] and pool [$F_{1,63} = 11.47, P < 0.001$] and treatment $\times$ pool [$F_{1,63} = 8.49, P < 0.005$] and disposition $\times$ pool interactions [$F_{1,63} = 4.94, P < 0.030$]. Compared with placebo, oxytocin increased in-group love among both cooperators [$t(23) = 2.98, P < 0.008$] and noncooperators [$t(40) = 3.81, P < 0.001$]. Oxytocin did not affect out-group hate [for cooperators, $t(23) = 0.96, P < 0.35$; for noncooperators, $t(40) = -0.11, P < 0.92$] (Fig. 2A). Because no interactions involving both treatment and disposition were significant [all $F_{1,63} < 0.48$, all $P > 0.40$], treatment and

disposition have parallel but independent effects on parochial altruism. This conclusion also applied to in-group trust and out-group distrust (Fig. 2B). Specifically, submitting in-group trust and out-group distrust to a 2 (treatment: oxytocin/placebo) $\times$ 2 (disposition: cooperators/noncooperators) $\times$ 2 (focus: in-group trust/out-group distrust) ANOVA showed no effects for disposition [all $F_{1,63} < 1$, all $P > 0.29$] but replicated the effect for focus [$F_{1,63} = 6.62, P < 0.012$] and the treatment $\times$ focus interaction [$F_{1,63} = 3.27, P < 0.075$]. Compared with placebo, oxytocin increased in-group trust [$M = 3.43$ versus $M = 4.51, t(65) = -3.19, P < 0.01$] but did not influence out-group distrust [$M = 3.30$ versus $M = 2.82, t(65) = 1.08, P < 0.28$].

In the IPD-MD, contributions to the between-group pool reflect aggression oriented toward downgrading the well-being of the out-group both absolutely and relative to the in-group. Oxytocin did not motivate such out-group hate. However, the IPD-MD provides in-group members with no direct means to protect (payoff to) the in-group against out-group aggression. It thus remains possible that in addition to in-group love, oxytocin modulates defensive aggression against

threatening out-groups. In Experiment 3, after intranasal administration of oxytocin or placebo, 75 males were randomly assigned to one of four different between-group prisoner dilemmas (BG-PD). Participants decided, on behalf of their in-group, to cooperate or not with another participant representing the out-group (26). The participant's choice to cooperate or not combined with the out-group's decision to cooperate or not yields four possible payoffs (Fig. 3A): temptation (T), reward (R), punishment (P), and sucker (S), which are ordered as $T > R > P > S$ (27). This ordering has several consequences. Figure 3B shows that if both participant and out-group cooperate, both in-group and out-group obtain the reward payoff of 1.00, which exceeds the punishment payoff for mutual noncooperation of 0.60 [i.e., $R - P = 1.00 - 0.60 = 0.40$ (henceforth cooperator's gain)]. The dilemma occurs because both the participant and the out-group representative obtain even higher payoffs for their own group by noncooperation. Noncooperation may reflect the greedy desire to exploit the out-group and/or the defensive desire to protect the in-group against out-group noncooperation. First, if the out-group were to cooperate, participants would obtain higher outcomes for their in-group by noncooperation (T) than by cooperation (R) [in

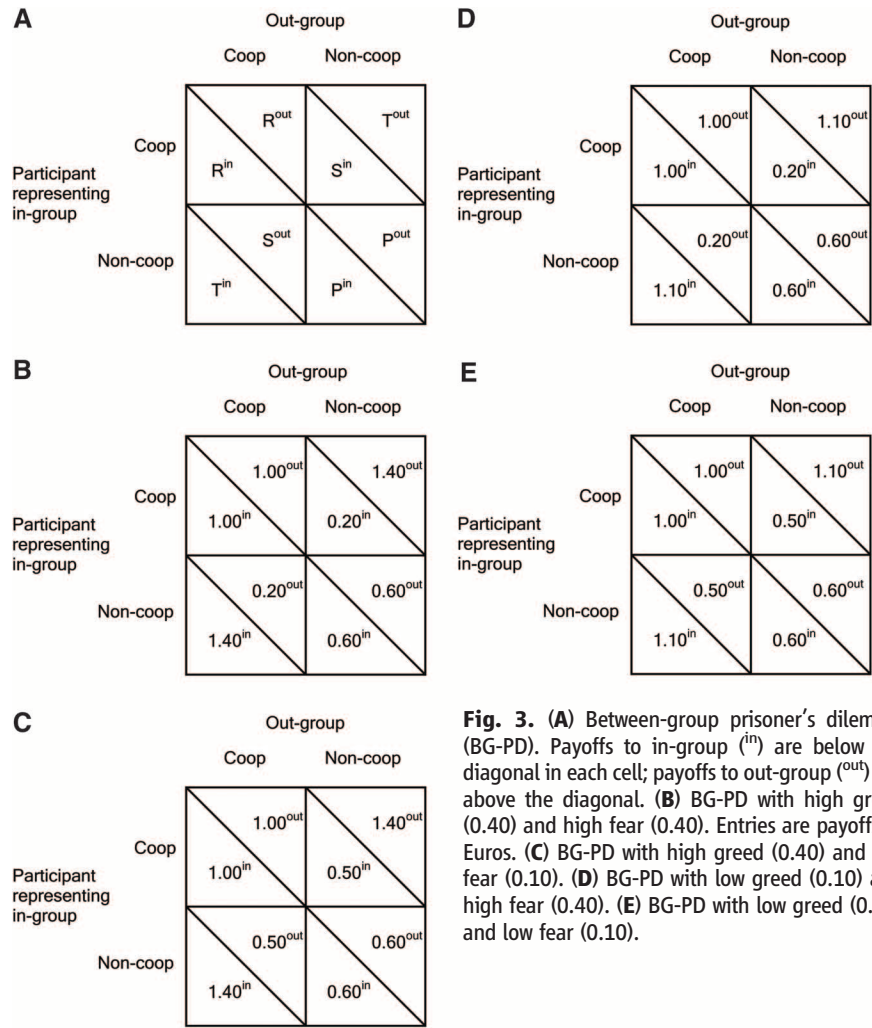


Fig. 3. (A) Between-group prisoner's dilemma (BG-PD). Payoffs to in-group (ⁱⁿ) are below the diagonal in each cell; payoffs to out-group (^{out}) are above the diagonal. (B) BG-PD with high greed (0.40) and high fear (0.40). Entries are payoffs in Euros. (C) BG-PD with high greed (0.40) and low fear (0.10). (D) BG-PD with low greed (0.10) and high fear (0.40). (E) BG-PD with low greed (0.10) and low fear (0.10).

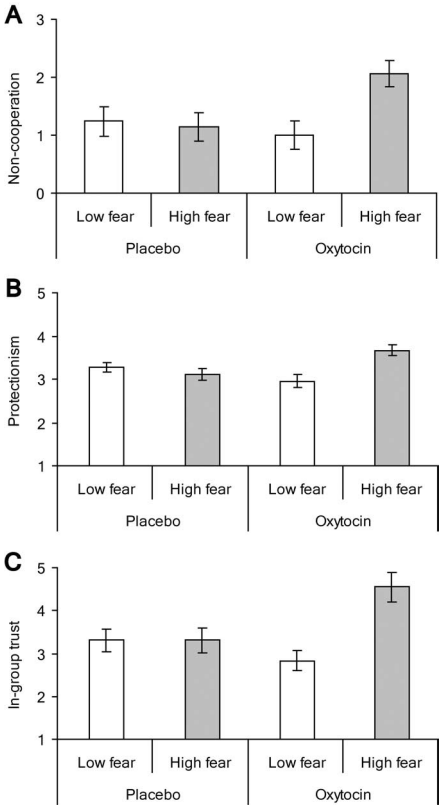


Fig. 4. (A) Noncooperation in the BG-PD (range 0 to 3, displayed $\pm$ SE). (B) Motivation to protect in-group members [ranging from 1 (low protectionism) to 5 (high protectionism), displayed $\pm$ SE]. (C) Participant's in-group trust [ranging from 1 (low) to 5 (high), displayed $\pm$ SE].

Fig. 3B: $T - R = 1.40 - 1.00 = 0.40$ (henceforth greed)]. Second, if the out-group were to non-cooperate, participants obtain higher outcomes for their in-group by noncooperation (P) than by cooperation (S) [in Fig. 3B: $P - S = 0.60 - 0.20 = 0.40$ (henceforth, fear)]. In Experiment 3, we manipulated payoffs so that the magnitude of greed and fear varied independently (28, 29). Greed was set at a high 0.40 in Fig. 3, B and C, and at a low 0.10 in Fig. 3, D and E. Because across these games cooperator's gain and fear are held constant, higher noncooperation in the games in Fig. 3, B and C, compared with those in Fig. 3, D and E, must reflect a greedy desire to exploit the out-group. Fear was set at a high 0.40 in Fig. 3, B and D, and at a low 0.10 in Fig. 3, C and E. Because cooperator's gain and greed are held constant, higher noncooperation in the games in Fig. 3, B and D, compared with those in Fig. 3, C and E, must reflect an anxious desire to protect the in-group against a possibly aggressive out-group.

Participants made three confidential decisions between cooperation and noncooperation (the choices by out-group members were not revealed) (23). Given that noncooperation promotes in-group welfare and hurts the out-group, and because oxytocin promotes in-group love, we hypothesized that oxytocin triggers noncooperation and explored whether this tendency would be stronger under high greed. More important, and following animal research (19, 20), we expected oxytocin to motivate protectionism: It promotes noncooperation especially when out-group fear is high. To test these hypotheses, noncooperation was submitted to a 2 (treatment: oxytocin/placebo) $\times$ 2 (greed: low/high) $\times$ 2 (fear: low/high) between-subjects ANOVA. More noncooperation was observed among individuals given oxytocin compared with placebo [$F_{1,67} = 3.98, P < 0.05$]. This effect was qualified by the treatment $\times$ fear interaction [$F_{1,67} = 7.51, P < 0.01$]: More noncooperation was seen among individuals given oxytocin rather than placebo when out-group fear was high rather than low (Fig. 4A). No effects involving greed were significant [all $F_{1,67} < 2.03, P > 0.16$]. If anything, the treatment $\times$ fear interaction was stronger under low greed. Overall, this shows that oxytocin promotes defense-motivated aggression in inter-group conflict.

To substantiate that noncooperation was related to the desire to protect one's in-group, participants rated whether they (i) tried to defend their in-group against possible out-group noncooperation (protectionism), (ii) expected fellow in-group members to serve in-group interests by noncooperation toward the out-group (in-group trust), and (iii) expected out-group noncooperation (out-group distrust) (23). A 2 (treatment) $\times$ 2 (greed) $\times$ 2 (fear) ANOVA showed that protectionism was stronger among individuals given oxytocin rather than placebo [$F_{1,67} = 4.27, P < 0.047$], especially when out-group fear was high [$F_{1,67} = 11.59, P < 0.001$]; no other effects were significant (Fig. 4B). Likewise, in-group trust was stronger among

individuals given oxytocin rather than placebo [$F_{1,67} = 8.19, P < 0.006$], especially when out-group fear was high [$F_{1,67} = 6.36, P < 0.014$]; no other effects were significant (Fig. 4C). Both protectionism and in-group trust correlated with participant noncooperation toward the out-group [Pearson $r(75) = 0.35, P < 0.002$, and $r(75) = 0.48, P < 0.001$, respectively]. Consistent with the previous experiments, a 2 (treatment) $\times$ 2 (greed) $\times$ 2 (fear) ANOVA on out-group distrust showed no effects [all $F_{1,67} < 1.22$, all $P > 0.28$ ($M = 5.73$)]. This shows that noncooperation toward the out-group was driven less by expectations about out-group aggression than by the extent to which such aggression would hurt the in-group. Put otherwise, these findings reflect a "tend and defend" pattern in which oxytocin stimulates humans to aggress against out-group threat in order to protect their in-group.

When in-groups face competing out-groups, humans are motivated to display parochial altruism, and such parochial altruism has strong survival functions. We have shown that specific forms of parochial altruism have their biological roots in oxytocin; across three experiments we found that compared with those given placebo, individuals given oxytocin displayed more in-group trust and in-group love but did not display more out-group hate and out-group distrust. We also showed that oxytocin (compared with placebo) led to defensive forms of out-group aggression when out-group threat was eminent. These findings generalized across dispositional cooperators and noncooperators. Recent work has linked such phenotypic differences with genetic differences in OXTR (12, 14), and new research may examine how genetic differences in OXTR relate to the regulation of intergroup conflict and competition. However, because our experiments involved males only, it is unknown whether oxytocin in females triggers the "tend and defend" form of parochial altruism uncovered here (30). This notwithstanding, because violent intergroup conflict more often involves males rather than females (the "male warrior hypothesis") (31), findings pertain to the most relevant half of the human species.

Others before us emphasized that parochial altruism contributes to individual survival (1, 10) and speculated that the human brain evolved to maintain and promote social life and to protect against eminent threats, including competing out-groups. Merging insights and techniques from the biological, economic, and psychological sciences, we uncovered a biological cause of intergroup competition and conflict. Our findings show that oxytocin, a neuropeptide functioning as both a neurotransmitter and hormone, plays a critical role in driving in-group love and defensive (but not offensive) aggression toward out-groups. Perhaps offensive forms of out-group hate have their biological roots elsewhere, or perhaps these tendencies are primarily grounded in perceived in-group love and protectionism in competing out-groups. After all, if competing out-groups become strong

and powerful, they become a threat to the in-group, and this in and of itself not only motivates in-group members to display in-group love but also motivates protectionism and preemptive strike. As shown here, this "tend and defend" form of parochial altruism is precisely what oxytocin modulates.

References and Notes

1. S. Bowles, *Science* **324**, 1293 (2009).
2. S. T. Fiske, *Curr. Dir. Psychol. Sci.* **11**, 123 (2002).
3. T. R. Cohen, C. A. Insko, *Perspect. Psychol. Sci.* **3**, 87 (2008).
4. M. B. Brewer, *J. Soc. Issues* **55**, 429 (1999).
5. J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
6. H. Arrow, *Science* **318**, 581 (2007).
7. G. Bornstein, *Pers. Soc. Psychol. Rev.* **7**, 129 (2003).
8. L. Berkowitz, *Psychol. Bull.* **106**, 59 (1989).
9. D. J. Albert, M. L. Walsh, R. H. Jonik, *Neurosci. Biobehav. Rev.* **17**, 405 (1993).
10. C. Darwin, *The Descent of Man* (Appleton, New York, 1873).
11. Z. R. Donaldson, L. J. Young, *Science* **322**, 900 (2008).
12. S. M. Rodrigues, L. R. Saslow, N. Garcia, O. P. John, D. Keltner, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21437 (2009).
13. P. J. Zak, R. Kurzban, W. T. Matzner, *Horm. Behav.* **48**, 522 (2005).
14. S. Israel et al., *PLoS ONE* **4**, e5535 (2009).
15. M. Heinrichs, B. von Dawans, G. Domes, *Front. Neuroendocrinol.* **30**, 548 (2009).
16. T. Baumgartner, M. Heinrichs, A. Vonlanthen, U. Fischbacher, E. Fehr, *Neuron* **58**, 639 (2008).
17. M. Kosfeld, M. Heinrichs, P. J. Zak, U. Fischbacher, E. Fehr, *Nature* **435**, 673 (2005).
18. J. A. Bartz, E. Hollander, *Horm. Behav.* **50**, 518 (2006).
19. O. J. Bosch, S. L. Meddle, D. I. Beiderbeck, A. J. Douglas, I. D. Neumann, *J. Neurosci.* **25**, 6807 (2005).
20. A. Campbell, *Biol. Psychol.* **77**, 1 (2008).
21. C. A. Pedersen, *Ann. N.Y. Acad. Sci.* **1036**, 106 (2004).
22. S. G. Shamay-Tsoory et al., *Biol. Psychiatry* **66**, 864 (2009).
23. Materials and methods are available as supporting material on Science Online.
24. N. Halevy, G. Bornstein, L. Sagiv, *Psychol. Sci.* **19**, 405 (2008).
25. P. A. M. Van Lange, *J. Pers. Soc. Psychol.* **77**, 337 (1999).
26. T. Wildschut, B. Pinter, J. L. Vevea, C. A. Insko, J. Schopler, *Psychol. Bull.* **129**, 698 (2003).
27. A. Rapoport, A. M. Chammah, *Prisoner's Dilemma: A Study in Conflict and Cooperation* (Univ. of Michigan Press, Ann Arbor, MI, 1965).
28. T. K. Ahn, E. Ostrom, D. Schmidt, R. Shupp, J. Walker, *Public Choice* **106**, 137 (2001).
29. S. S. Komorita, C. D. Parks, *Annu. Rev. Psychol.* **46**, 183 (1995).
30. G. Domes et al., *Psychoneuroendocrinology* **35**, 83 (2010).
31. M. Van Vugt, *Ann. N.Y. Acad. Sci.* **1167**, 124 (2009).
32. The authors thank H. Aaldering, H. Ter Kuilen, N. Halevy, and M. Kemper for their support at various stages of the project. This research was supported by grant NWO 400-06-098 of the Netherlands Science Foundation to Carsten K. W. De Dreu. The authors declare no competing interests.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1408/DC1
Materials and Methods
References

3 March 2010; accepted 28 April 2010
10.1126/science.1189047

Focal Lesions of Human Hippocampal CA1 Neurons in Transient Global Amnesia Impair Place Memory

T. Bartsch,^{1*} R. Schönfeld,² F. J. Müller,³ K. Alfke,⁴ B. Leplow,² J. Aldenhoff,³ G. Deuschl,¹ J. M. Koch³

A critical role in place learning has been attributed to place cells within the cornu ammonis 1 (CA1) sector of the hippocampus in rodents. The role of CA1 cells in the human hippocampus with regard to place learning remains elusive. Using a virtual Morris water maze, we investigated patients with acute transient global amnesia (TGA), a rare self-limiting dysfunction of the hippocampal system. Fourteen individuals with selective and focal lesions in the CA1 sector of the hippocampus showed a profound impairment in place learning. The size of the lesions and the duration of the TGA correlated with the deficit in the performance.

The hippocampus is a core structure for the processing of topographical memory and place learning tasks. Its functions include the processing of spatial information in terms of a representation of an allocentric (world-centered) spatial position within a given environment (*1*). The role of hippocampal cornu ammonis 1 (CA1) neurons has been of particular interest to researchers, because data from rodents and primates show that hippocampal CA1 neurons are critically involved in the process of navigation and the consolidation of spatial memory (*2–4*). Single-unit recordings from hippocampal CA1 pyramidal neurons in rodents show that spatial information is encoded in the firing pattern of CA1 and CA3 neurons (*5*). Hippocampal lesions in animals are sufficient to produce a substantial impairment in spatial memory and a disruption of the ability to acquire or use a spatial mapping strategy or representation (*6–9*).

Complete understanding of the contribution of CA1 neurons in the human hippocampus to spatial memory remains elusive because of the lack of a human model with spatial resolution good enough to analyze the contributions of hippocampal subfields to declarative memory (*10, 11*). Hence, a selective and focal disruption of neuronal elements of hippocampal circuits might allow deeper insights into the functional anatomy and processing of spatial information in humans.

Transient global amnesia (TGA) is a rare amnesic syndrome that is characterized by a sudden onset of a selective antero- and retrograde amnesia without any further neurological deficits (*12, 13*). The time course of the acute amnesic syndrome is, by clinical definition, limited to up

to 24 hours but usually lasts 6 to 10 hours. In patients with TGA, highly focal lesions confined to the CA1 field of the hippocampal cornu ammonis can be detected with high-resolution magnetic resonance imaging (MRI) 24 to 72 hours after the amnesic phase (*14, 15*). Typically, lesions outside the hippocampus are absent. These MRI-detected lesions are the structural correlate of the amnesic deficit, reflecting a focal transient perturbation of CA1-dependent hippocampal circuits (*16, 17*). Lesions defined by diffusion-weighted imaging (DWI) are accompanied by corresponding lesions in T2-weighted images, and both are reversible within 14 days after TGA. Structural as well as functional sequelae 4 to 6 months after TGA onset are missing (*18, 19*).

A total of 14 patients [8 women, 6 men, mean age 67.8 ± 7.6 years (mean $\pm$ SD)] presenting with TGA episodes were included in the study and compared with 10 healthy controls (5 women, 5 men, mean age 67.0 ± 7.1 years) (table S1). All patients studied with MRI presented hippocampal lesions in a time window of 48 to 72 hours after onset. The size of focal hyperintense lesions ranged from 1 to 7 mm² and corresponded to DWI and T2 signal changes (Fig. 1 and fig. S2). The average size of CA1 lesions in their maximal extension in the coronal plane was 6.4 ± 4.7 mm² (right side, 5.9 ± 4.8 mm²; left side, 7.1 ± 4.7 mm²). Lesions were distributed along the anterior-posterior axis within the hippocampus and showed an edema-like configuration in the T2-weighted images (Fig. 1 and fig. S3). A detailed analysis of the distribution of lesions showed that all lesions were selectively found in the area corresponding to the CA1 sector (Sommer sector) of the hippocampal cornu ammonis (figs. S2 and S3). A detailed whole-brain study did not show evidence for diffusion lesions outside the hippocampal cornu ammonis (fig. S2).

In our virtual training task, there were no significant group differences between TGA patients and controls [latencies: controls, 3.45 ± 1.34 min; TGA, 4.23 ± 2.23 min; $t(19) = 0.84$; $P = 0.486$; path length: controls, 1.54 ± 0.47 ; TGA, 2.15 ± 1.48 ; $t(19) = 1.23$; $P = 0.234$]. TGA patients

coped with our virtual training environment as well as controls did.

In the cued-target exploration, controls and patients did not differ in the latency measure [$t(21) = 1.51$, $P = 0.147$]. Group means of the path length measure were 0.91 ± 0.06 for the controls and 1.36 ± 0.73 for TGA patients. Both groups showed a large and significant distortion of group variances in the path length measure [$F_{1,21} = 6.43$, $P = 0.019$]. Significant differences in group means were detected in the path length measure (exact P value = 0.047), with TGA patients taking longer paths than controls did. After contact with the hidden target, patients and controls showed no differences in inspecting the island [latencies: controls, $1:22 \pm 0:36$ min; TGA, $1:07 \pm 0:36$ min; $t(21) = 1.04$; $P = 0.308$; path length: controls, 0.74 ± 0.52 ; TGA, 0.87 ± 0.64 ; $t(21) = 0.53$; $P = 0.604$] (Fig. 2A).

We observed a large and significant impairment in acute TGA patients during the acquisition trials where spatial learning should occur. Main effects for the latency [$F_{1,21} = 23.05$, $P < 0.002$] and the path length [$F_{1,21} = 35.50$, $P < 0.001$] were significantly different, indicating that TGA patients took more time and longer paths to navigate to the hidden target (Fig. 2A).

No main effect for the within-subject factor [blocks of acquisition trials, see supporting online material (SOM) text] and no interaction with TGA status (TGA versus controls) were detected. Hence, all acquisition trials were conflated into a single level and, together with the cued-target exploration trial, subjected to a further repeated measures analysis of variance (ANOVA), which was similar to the within-subject factor. The TGA status by trial interaction reached statistical significance for the latency [$F_{1,21} = 11.42$, $P < 0.002$] and the path length measure [$F_{1,21} = 17.33$, $P < 0.001$]. In the path length measure, the group differences increased from a significant effect with $d = 0.8$ to 1.6 (Cohen's d). For TGA patients compared with controls, this indicates a large decrease in the performance from the cued-target exploration to the acquisition trials. TGA patients showed a normal behavior in the cued-target trial, but they exhibited no systematic search behavior in the acquisition trials (Fig. 3).

Three different probe trials were performed at follow-up when acute symptoms had completely resolved in order to measure recall performance (fig. S1). Behavioral data of the first and second probe trial show TGA patients with profound deficits in completing the virtual water maze (VWM) task (Fig. 2B). We found large and significant group differences in both trials in the latency measure [$t(19) = 3.69$, $P = 0.002$; $t(20) = 2.80$, $P = 0.011$] as well as in the path length measure [$t(19) = 3.69$, $P = 0.002$; $t(20) = 3.56$, $P = 0.002$]. The effect sizes of group differences ranged from $d = 1.1$ to 1.3.

The third probe trial was a target reversal condition where the measure of the relative dwell time for the original quadrant showed profound deficits in the TGA patients (Fig. 4A). Patients

¹Department of Neurology, University Hospital Schleswig-Holstein, University of Kiel, Kiel, Germany. ²Institute of Psychology, Martin-Luther-University Halle-Wittenberg, Halle, Germany. ³Department of Psychiatry, University Hospital Schleswig-Holstein, University of Kiel, Kiel, Germany. ⁴Institute of Neuroradiology, University Hospital Schleswig-Holstein, University of Kiel, Kiel, Germany.

*To whom correspondence should be addressed. E-mail: t.bartsch@neurologie.uni-kiel.de

spent significantly less time in the former target quadrant [$t(20) = 4.31, P < 0.001, d = -1.4$]. TGA patients searched the VWM quadrants randomly (Fig. 4B). For the mean initial heading error, patients significantly differed from controls [$t(20) = 3.10, P = 0.006$]. The mean difference was 15° on average and resulted in a large effect ($d = 1.1$). No

differences were found in the path length measure, which we attributed to a comparable general search activity in both groups during this VWM trial.

The degree of the severity of the TGA deficit was compared with lesion size and TGA duration. No substantial correlation was found

between both variables, but each variable correlated significantly with measures of verbal and spatial learning behavior in the acute phase of TGA. Lesion size was associated with the Rey complex figure (RCF) recall performance ($r = -0.52, P = 0.067, n = 13$ patients). TGA duration was correlated with RCF recall ($r = -0.62, P = 0.024, n = 13$) and the average path length traveled during the VWM acquisition trials ($r = 0.58, P = 0.048, n = 12$). This indicates that a lower learning performance was associated with the severity of the TGA. The path length and the RCF recall showed a significant intercorrelation ($r = -0.59, P = 0.043, n = 12$) (20).

TGA severity was also significantly associated with the VWM recognition performance after 14 days. Lesion size correlated with latency ($r = 0.64, P = 0.026, n = 12$) and path length ($r = 0.68, P = 0.014, n = 12$) measured in the first probe trial. Similarly, TGA duration was correlated with the latency and path length of the first probe trial (path length: $r = 0.62, P = 0.030, n = 12$; latency: $r = 0.61, P = 0.036, n = 12$) and the path length of the second probe trial ($r = 0.64, P = 0.025, n = 12$).

We conducted a regression analysis to predict the behavioral response in the VWM recognition trial. Because the relationship between path length and lesion size as well as TGA duration appeared to be nonlinear (Fig. 2C), a logarithmical fit was done with the regression model. Lesion size predicted 58.5% [$F_{1,11} = 14.90, P < 0.004$] of the variance in the path length measure. TGA duration proved to also be a significant predictor of recognition performance and contributed 47.2% [$F_{1,11} = 8.94, P < 0.014$] of the variance. After their symptoms had resolved, patients with

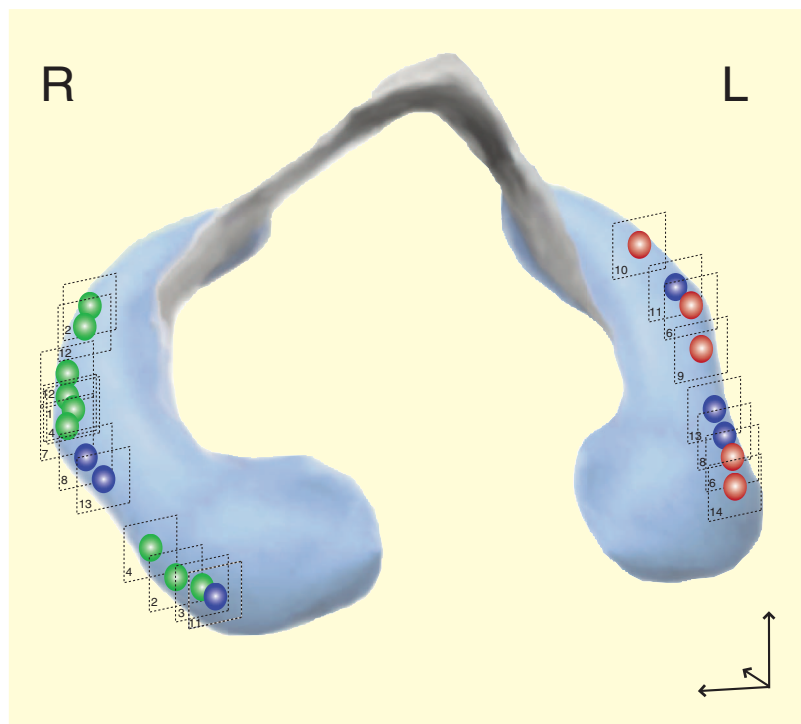


Fig. 1. Three-dimensional model of the hippocampus showing the location of lesions in the lateral hippocampus and the distribution along the anterior-posterior axis of the hippocampus. Red circles indicate left (L) lesions, green circles show lesions in the right (R) hippocampus, and blue circles indicate bilateral lesions. [Template modified after the Digital Anatomist Project (www9.biostr.washington.edu/da.html) and used with permission]

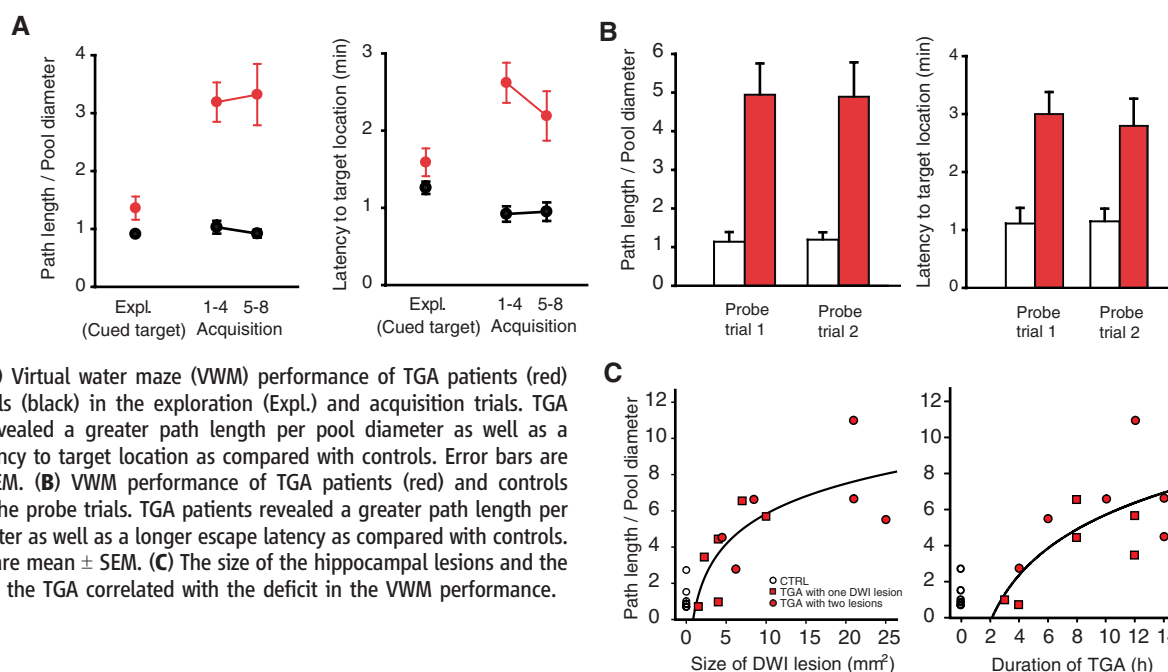


Fig. 2. (A) Virtual water maze (VWM) performance of TGA patients (red) and controls (black) in the exploration (Expl.) and acquisition trials. TGA patients revealed a greater path length per pool diameter as well as a longer latency to target location as compared with controls. Error bars are mean $\pm$ SEM. (B) VWM performance of TGA patients (red) and controls (white) in the probe trials. TGA patients revealed a greater path length per pool diameter as well as a longer escape latency as compared with controls. Error bars are mean $\pm$ SEM. (C) The size of the hippocampal lesions and the duration of the TGA correlated with the deficit in the VWM performance.

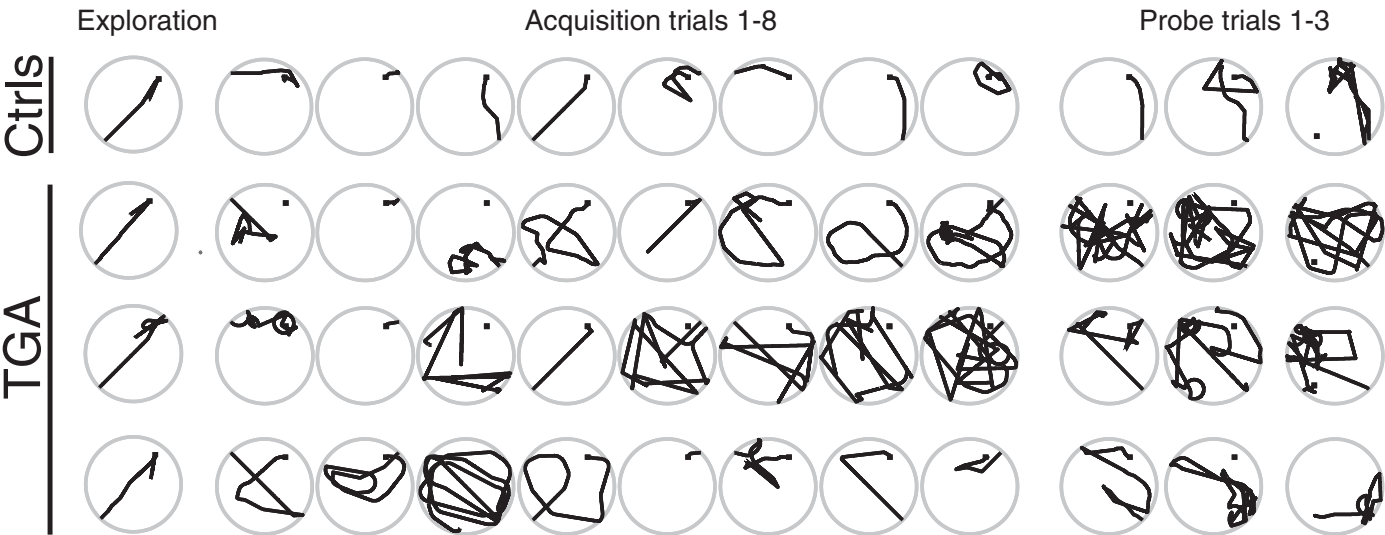


Fig. 3. Representative raw data of the VWM search paths of a control person (Ctrls) and three TGA patients.

smaller DWI lesions and those with a shorter TGA interval traveled shorter distances to locate the hidden target.

The testing in the acute episode revealed substantial memory deficits in declarative verbal and visuo-constructive memory subtests [Rey auditory verbal learning test (RAVLT) and RCF] in 14 patients as compared with controls (table S2). Patients showed profound floor effects in the RAVLT retention and delayed recognition task. At the time of follow-up, patients did significantly improve in reaching the control reference range (table S2). The follow-up testing was performed after patients were fully recovered from all TGA symptoms after 14 days (controls, 14 ± 1.2 days; TGA patients, 14 ± 4.5 days). Patients did not differ from controls in naming and conceptual knowledge, premorbid general intellectual abilities, and visual attention in the follow-up testing (table S3).

Our findings show that acute and focal lesions of CA1 neurons in the human cornu ammonis lead to a profound impairment of place learning. The validity of a virtual reality approach, including the modified Morris water maze used here, in testing spatial memory function in young and aged patients has been shown in functional imaging and lesion studies (21, 22).

The selectivity of CA1 lesions is critical to the interpretation of the present data. The focal hyperintense MRI-identified lesions, indicating restricted diffusion that emerges in the course of TGA, were selectively confined to the CA1 sector of the cornu ammonis, so that TGA is a natural lesion model of a perturbation of CA1 neurons (14, 16, 23). The functional impairment of CA1 lesions was based on the detection of hyperintense diffusion lesions that sensitively mirror an impaired cellular metabolism. The anatomical mapping within the different subfields of the cornu ammonis was based on a very conservative approach in terms of identifying and correlating the DWI lesions to high-resolution T2-weighted images in different planes.

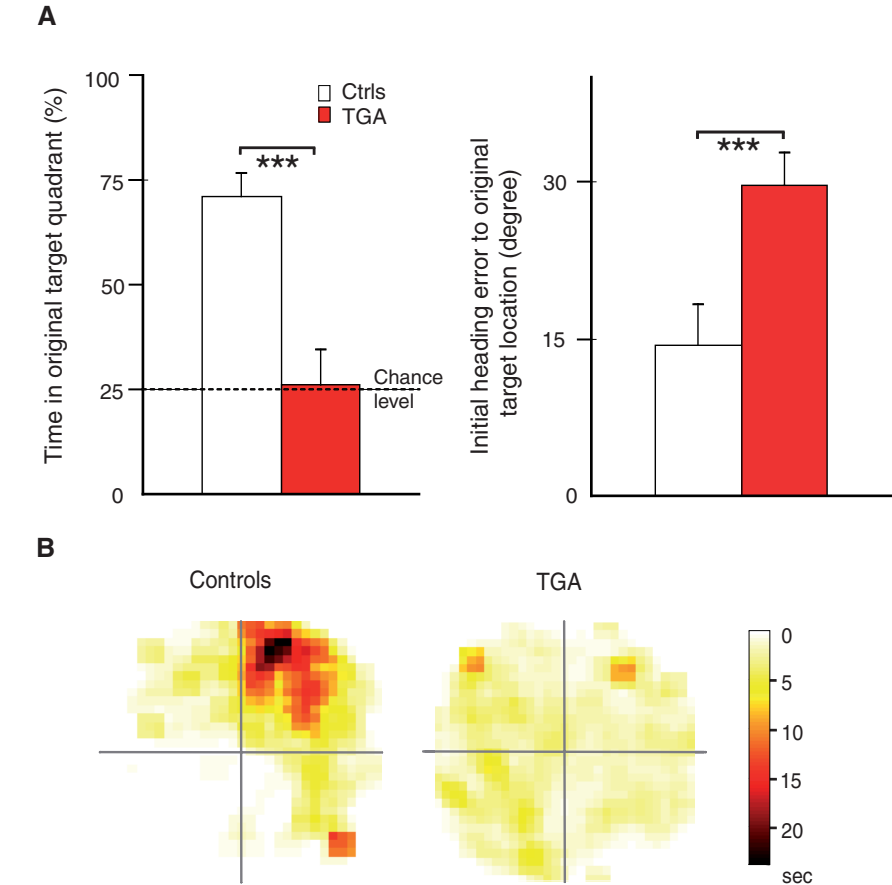


Fig. 4. (A) Results from the third probe trial: mean relative dwell time in the original target quadrant and mean initial heading error to original target location. Patients (red) spent significantly less time in the learned target quadrant than controls (white) did. Error bars are mean $\pm$ SEM. *** $P < 0.01$. (B) Density plots (heat map) for grouped data showing where controls and TGA patients concentrated their search. Dark red colors indicate areas where more than 20 s were spent. Controls showed a clear preference for the original target location, whereas patients appeared to search the maze randomly.

No diffusion lesions outside the CA1 fields were detected, and this makes it unlikely that covert lesions went undetected. Moreover, TGA patients were tested within a few hours after the onset of the amnesic syndrome, when the patients' deficit was most extensive, with a retention span of a few minutes. Considering this temporal course, the amnesic deficit mirrors the acute perturbation of

CA1-dependent network properties without a compensatory reorganization of memory circuits.

The nature of the observed MRI changes and hence the pathophysiological mechanisms during TGA are still unclear. Neurons in the CA1 sector of the hippocampal cornu ammonis are characterized by a selective vulnerability to metabolic and oxidative stress induced by hypoxemia, β -amyloid neurotoxicity, and ischemia that is mediated by cellular glutamate overload and subsequent calcium influx (24). Several pathogenic mechanisms have been suggested, such as migraine-related mechanisms, cortical spreading depression, ischemia, venous flow abnormalities, and seizure activity that may lead to impairment of cellular diffusion with subsequent DWI changes (13). Focal MR spectroscopy of DWI/T2 lesions in patients with TGA revealed a lactate peak in TGA patients showing a hippocampal lesion (17). This suggests that anaerobic glycolysis as a cellular stress response of CA1 neurons is the metabolic correlate of the diffusion changes seen in TGA.

The understanding of the functional role of the hippocampus in the processing of memory in humans has been greatly advanced by studying the amnesic deficit of patients with lesions in the hippocampus (25, 26). Partial or extended lesions to the hippocampus lead to a global amnesic syndrome covering various memory domains, including spatial memory (25, 27, 28). The functional role of CA1 neurons in spatial memory has been extensively studied in rodents. In primates and humans, however, the functional role is less clear, because there are only a limited number of studies, and results are often ambiguous (29). In rodent and primate models aimed at selective lesioning of CA1 neurons by using a transient ischemia or neurochemical lesioning, animals showed an impairment in the establishment of allocentric spatial representations, spatial navigation, and spatial relational learning, dependent on the paradigm used, although the overall impairment was frequently mild and the performance partly superior to that of controls (29–31). Focal hippocampal lesions in monkeys resulted in greater memory deficits than did surgical resection of the hippocampal formation (32).

Apart from the involvement of the hippocampus in spatial navigation, neuroimaging studies support the idea that a widespread hippocampal-neocortical network has a role in the processing of spatial navigation in humans. Recent neuroimaging findings using a virtual-reality task suggest a dissociation between the retrosplenial cortex and the hippocampus, inasmuch as the activation of the retrosplenial cortex correlated with the integration of egocentric spatial information with cues about self-motion (33, 34). In contrast, hippocampal activity reflected the degree of learning in the navigational task given. Although these findings support the notion of a network processing of spatial memory tasks, the dynamics and interactions are not fully understood (35). Our results highlight the critical relay function of hippocampal CA1 neurons within this hippocampal-parietal network.

References and Notes

1. E. I. Moser, E. Kropff, M. B. Moser, *Annu. Rev. Neurosci.* **31**, 69 (2008).
2. J. O'Keefe, J. Dostrovsky, *Brain Res.* **34**, 171 (1971).
3. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, Oxford, 1978); available at www.cognitivemap.net/.
4. N. J. Broadbent, L. R. Squire, R. E. Clark, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14515 (2004).
5. B. L. McNaughton *et al.*, *J. Exp. Biol.* **199**, 173 (1996).
6. R. G. Morris, P. Garrud, J. N. Rawlins, J. O'Keefe, *Nature* **297**, 681 (1982).
7. E. Moser, M. B. Moser, P. Andersen, *J. Neurosci.* **13**, 3916 (1993).
8. R. E. Clark, S. M. Zola, L. R. Squire, *J. Neurosci.* **20**, 8853 (2000).
9. A. L. Griffin, H. Eichenbaum, M. E. Hasselmo, *J. Neurosci.* **27**, 2416 (2007).
10. H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, H. Tanila, *Neuron* **23**, 209 (1999).
11. N. Burgess, E. A. Maguire, J. O'Keefe, *Neuron* **35**, 625 (2002).
12. J. R. Hodges, C. P. Warlow, *J. Neurol. Neurosurg. Psychiatry* **53**, 834 (1990).
13. T. Bartsch, G. Deuschl, *Lancet Neurol.* **9**, 205 (2010).
14. O. Sedlacek *et al.*, *Neurology* **62**, 2165 (2004).
15. H. Y. Lee *et al.*, *Neuroradiology* **49**, 481 (2007).
16. T. Bartsch *et al.*, *Brain* **129**, 2874 (2006).
17. T. Bartsch *et al.*, *Neurology* **70**, 1030 (2008).
18. A. Cianfoni *et al.*, *Arch. Neurol.* **62**, 1468 (2005).
19. T. Bartsch, K. Alfke, G. Deuschl, O. Jansen, *Ann. Neurol.* **62**, 475 (2007).
20. Because TGA patients showed a profound floor effect in the RAVLT, we did not compute correlations with these measures (table S2).
21. E. A. Maguire, *Philos. Trans. R. Soc. London Ser. B* **352**, 1475 (1997).
22. R. Schoenfeld, N. Moenich, F. J. Mueller, W. Lehmann, B. Leprow, *Behav. Brain Res.* **208**, 169 (2010).
23. Y. C. Weon, J. H. Kim, J. S. Lee, S. Y. Kim, *Am. J. Neuroradiol.* **29**, 1324 (2008).
24. X. Wang *et al.*, *BMC Neurosci.* **10**, 12 (2009).
25. S. Zola-Morgan, L. R. Squire, D. G. Amaral, *J. Neurosci.* **6**, 2950 (1986).
26. H. J. Spiers, E. A. Maguire, N. Burgess, *Neurocase* **7**, 357 (2001).
27. S. Abrahams, A. Pickering, C. E. Polkey, R. G. Morris, *Neuropsychologia* **35**, 11 (1997).
28. T. Hartley, E. A. Maguire, H. J. Spiers, N. Burgess, *Neuron* **37**, 877 (2003).
29. P. B. Lavenex, D. G. Amaral, P. Lavenex, *J. Neurosci.* **26**, 4546 (2006).
30. R. N. Auer, M. L. Jensen, I. Q. Whishaw, *J. Neurosci.* **9**, 1641 (1989).
31. I. Q. Whishaw, M. R. Rod, R. N. Auer, *Brain Res. Bull.* **34**, 283 (1994).
32. J. Bachevalier, M. Mishkin, *Neuropsychologia* **27**, 83 (1989).
33. T. Wolbers, C. Büchel, *J. Neurosci.* **25**, 3333 (2005).
34. R. A. Epstein, *Trends Cogn. Sci.* **12**, 388 (2008).
35. C. M. Bird, N. Burgess, *Nat. Rev. Neurosci.* **9**, 182 (2008).
36. We thank O. Jansen for helpful input and support and R. Rosenträger for technical assistance. This study has been supported by the Deutsche Forschungsgemeinschaft [to T.B., grant SFB 654 (TP A14); to J.M.K., grant KO 2067/1-2]; the Faculty of Medicine, University of Kiel, Germany (to T.B. and F.J.M.); and the Else Kröner-Fresenius-Stiftung (to F.J.M.). K.A. is currently affiliated with the Institute of Radiology and Neuroradiology, HELIOS Kliniken, Schwerin, Germany.

Supporting Online Material

www.sciencemag.org/cgi/content/full/328/5984/1412/DC1

Materials and Methods

SOM Text

Figs. S1 to S3

Tables S1 to S3

References

10 February 2010; accepted 22 April 2010

10.1126/science.1188160

Dlg1-PTEN Interaction Regulates Myelin Thickness to Prevent Damaging Peripheral Nerve Overmyelination

Laurent Cotter,^{1*} Murat Özçelik,^{1*†} Claire Jacob,¹ Jorge A. Pereira,¹ Veronica Locher,¹ Reto Baumann,¹ João B. Relvas,^{1,2} Ueli Suter,¹ Nicolas Tricaud^{1‡}

The thickness of the myelin sheath that insulates axons is fitted for optimal nerve conduction velocity. Here, we show that, in Schwann cells, mammalian disks large homolog 1 (Dlg1) interacts with PTEN (phosphatase and tensin homolog deleted on chromosome 10) to inhibit axonal stimulation of myelination. This mechanism limits myelin sheath thickness and prevents overmyelination in mouse sciatic nerves. Removing this brake results also in myelin outfoldings and demyelination, characteristics of some peripheral neuropathies. Indeed, the Dlg1 brake is no longer functional in a mouse model of Charcot-Marie-Tooth disease. Therefore, negative regulation of myelination appears to be essential for optimization of nerve conduction velocity and myelin maintenance.

The thickness of the myelin sheath is an essential parameter governing nerve conduction velocity (NCV) (1). This velocity falls from its optimum when myelin is too thick or too thin. Optimal conduction is obtained by adjusting the length and thickness of the myelin sheath to the axon diameter (2). To achieve this in peripheral nerves, myelinated axons express neuregulin-1 type III, which stimulates myelination by Schwann cells (SCs) (3, 4). Myelination slows down after the first two postnatal weeks

in mice as optimal thickness is approached (5), which suggests regulatory mechanisms that curb myelination.

We have previously characterized a mouse model of Charcot-Marie-Tooth disease 4B (CMT4B), a peripheral neuropathy caused by mutations in *MTMR2* (6). Schwann cell hypermyelination with focally folded myelin is the morphological disease hallmark (7), which indicates dysregulation of myelin production. *MTMR2* interacts with mammalian disks large homolog

1 (Dlg1, also SAP-97) (8), a PDZ-containing membrane-associated guanylate kinase (MAGUK)-family protein involved in basic cellular events including cell polarization, protein trafficking, and tumorigenesis (9). Here, we describe Dlg1 function during SC myelin sheath formation.

¹Institute of Cell Biology, Department of Biology, Eidgenössische Technische Hochschule (ETH) Zürich, CH-8093 Zürich, Switzerland.

²Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal.

*These authors contributed equally to this work.

†Present address: McKinsey & Company, Herrngasse 1-3, A-1010 Vienna, Austria.

‡To whom correspondence should be addressed. E-mail: nicolas.tricaud@cell.biol.ethz.ch

Dlg1 expression was silenced in SCs cocultured with DRG neurons (10) (fig. S1A). We confirmed that Dlg1-silenced SCs failed to myelinate (fig. S1B) (11). These cells also showed migration defects (fig. S1, C and D) and expressed less polarity protein Par3 (fig. S1E), which is required for cell migration (12) and myelination (13). Occasionally, silenced SCs overcame their migration defect and myelinated (fig. S1F). The resulting myelin segments were thicker than those of controls (fig. S1G), which indicated a role for Dlg1 in regulating myelin sheath thickness.

We then silenced Dlg1 expression efficiently in myelinating SCs in mouse sciatic nerves at early stages of myelination (3 to 4 days postnatal,

dPN) by injecting lentiviral vectors expressing short hairpin RNAs (shRNAs) (fig. S2) (10, 14). Two months after injection, we found a strikingly increased thickness of myelinating SCs silenced for Dlg1 compared with control cells (Fig. 1, A and B; $9.3 \pm 0.2 \mu\text{m}$ versus $7.1 \pm 0.3 \mu\text{m}$, $P < 0.001$). Average cell lengths remained unaffected (Fig. 1B, $347 \pm 11 \mu\text{m}$ versus $363 \pm 19 \mu\text{m}$, $P = 0.49$). The effects were SC autonomous, because Dlg1-silenced cells were thicker than neighboring cells on the same axon (Fig. 1C).

Coexpression of placental alkaline phosphatase (PLAP) allowed detection of infected SCs by electron microscopy (Fig. 1D). Compact myelin of Dlg1-silenced cells appeared structurally nor-

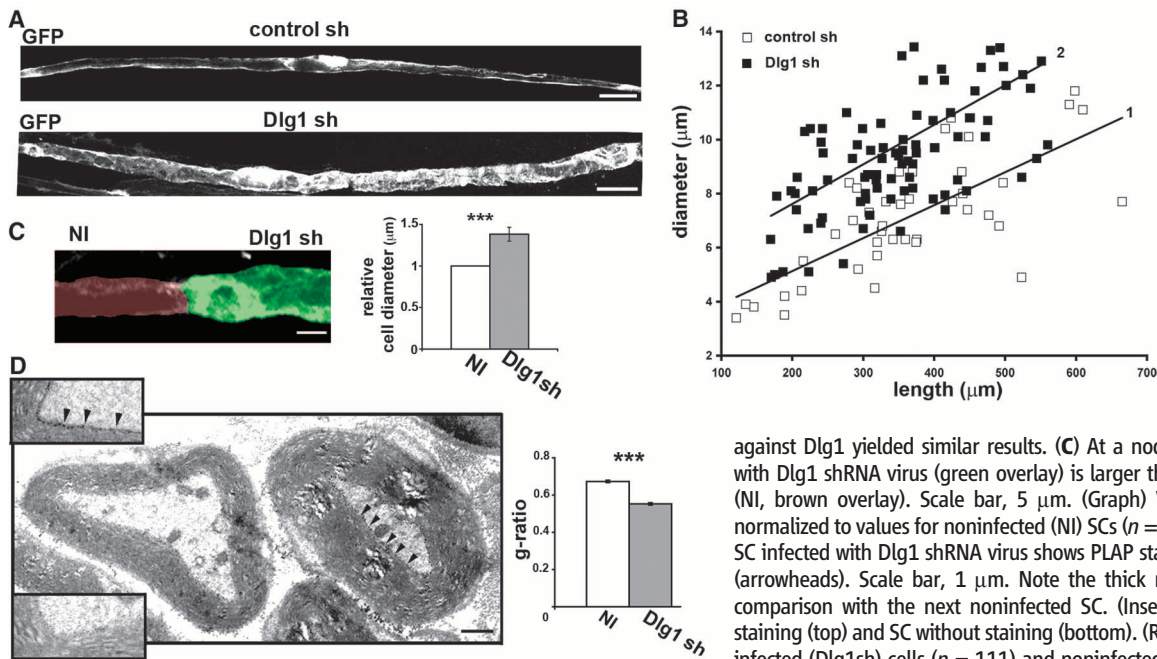
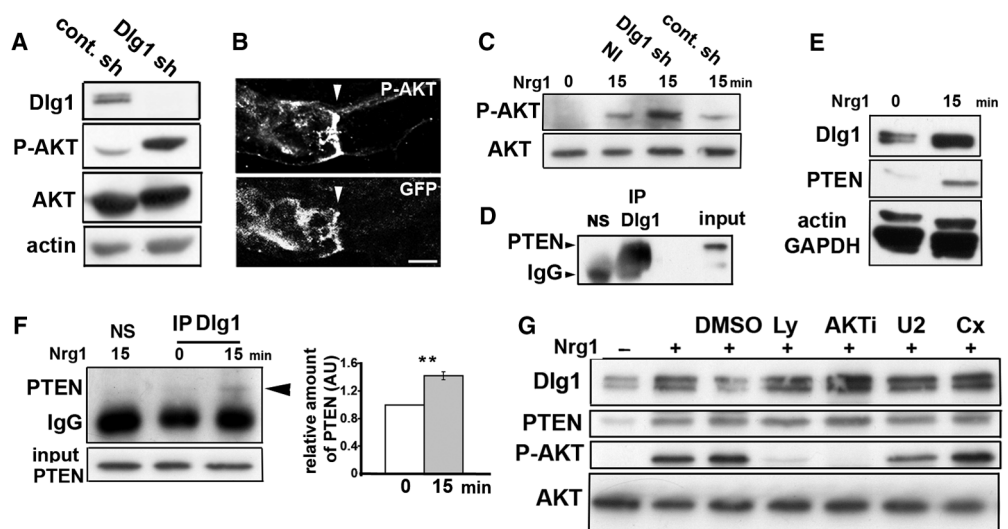


Fig. 1. Dlg1 silencing increases myelin sheath thickness. (A) SC infected with Dlg1 shRNA virus (bottom) shows a larger mean diameter than SC infected with control shRNA virus (top). Scale bars, 20 μm . (B) Diameters versus lengths of SCs infected with Dlg1 or control shRNAs viruses. Lines: Linear regressions for control ($n = 43$), slope line 1 and Dlg1 ($n = 87$), line 2 shRNAs. Two independent shRNAs

against Dlg1 yielded similar results. (C) At a node of Ranvier, a SC infected with Dlg1 shRNA virus (green overlay) is larger than the next noninfected SC (NI, brown overlay). Scale bar, 5 μm . (Graph) Values for infected SCs are normalized to values for noninfected (NI) SCs ($n = 13$ nodes) ($\pm\text{SEM}$). (D) (Left) SC infected with Dlg1 shRNA virus shows PLAP staining at interface with axon (arrowheads). Scale bar, 1 μm . Note the thick myelin sheath of this SC in comparison with the next noninfected SC. (Insets) Details of SC with PLAP staining (top) and SC without staining (bottom). (Right) Mean g ratio ($\pm\text{SEM}$) of infected (Dlg1sh) cells ($n = 111$) and noninfected (NI) cells ($n = 172$).

Fig. 2. Dlg1 inhibits AKT activation via interaction with PTEN. (A) Cultured SCs silenced for Dlg1 (Dlg1sh) have more phospho-AKT (P-AKT). For quantifications, see fig. S5A. (B) At a node of Ranvier (arrowhead), a myelinated SC infected with Dlg1 shRNA virus (GFP) has more P-AKT than the next noninfected SC. Scale bar, 5 μm . (C) Cultured Dlg1-silenced SCs (Dlg1sh) stimulated with neuregulin-1 (Nrg1) for 15 min generate more P-AKT than cells infected with control shRNA virus or noninfected cells (NI). Representative results of two independent experiments. (D) PTEN coimmunoprecipitates with Dlg1 (IP Dlg1) but not with nonimmune serum (NS) in rat sciatic nerves at 20 dPN. Representative results of two independent experiments. (E) Nrg1 stimulation of SCs increases Dlg1 and PTEN amounts after 15 min. Representative results of four independent experiments. (F) More PTEN coimmunoprecipitates with Dlg1 when SCs are stimulated with Nrg1. (Graph) Relative amount of coimmunoprecipitated PTEN in stimulated and nonstimulated SCs ($\pm\text{SEM}$; for four independent experiments). No PTEN is detected when nonimmune serum (NS) is used. (G) Cul-



tured SCs, left untreated or treated with Nrg1 for 15 min, and vehicle dimethylsulfoxide (DMSO) or inhibitors of PI3K (Ly), AKT (AKTi), MAPK (U2), and mRNA translation (cycloheximide, Cx). Representative results of three experiments.

mal (fig. S3), except that myelin was thicker (Fig. 1D). Consequently, the *g* ratio (axon diameter/total fiber diameter) was significantly reduced (Fig. 1D), with no change in average axon diameters ($10.7 \pm 0.4 \mu\text{m}$ versus $11 \pm 0.2 \mu\text{m}$, $P = 0.5$).

We also overexpressed Dlg1 fused with green fluorescent protein (GFP) in SCs of mouse sciatic nerves. Two months after birth, the majority of GFP-Dlg1-positive cells were non-myelin-forming cells; most cells expressing GFP alone had mye-

linated (fig. S4), which suggested that Dlg1 overexpression prevented SC myelination.

Our data indicated that Dlg1 acts as an inhibitor of SC myelination. Myelination is stimulated by axonal neuregulin-1 type III activating erbB2 and erbB3 receptors and stimulating the phosphatidylinositol 3-kinase (PI3K)-AKT pathway in SCs (15). Thus, we tested whether Dlg1 interferes with AKT activation. Indeed Dlg1 silencing induced a significant increase of AKT activation in cultured SCs (Fig. 2A and fig. S5A), as well as in myelinating SCs in vivo (Fig. 2B). Moreover, Dlg1 reduction potentiated AKT activation upon neuregulin-1 stimulation in cultured SCs (Fig. 2C), which demonstrated that Dlg1 inhibits neuregulin-driven AKT activation.

Dlg1 can interact directly with PTEN (phosphatase and tensin homolog deleted on chromosome 10) (16, 17), an inhibitor of AKT activation (18). This interaction may stabilize PTEN (19), which results in less AKT activation (17). We could coimmunoprecipitate PTEN with Dlg1 from both cultured SCs (fig. S5B) and sciatic nerves (Fig. 2D). Moreover, Dlg1 silencing reduced PTEN levels, whereas Dlg1 overexpression increased them (fig. S5C), which indicated that Dlg1 regulates PTEN amounts in SCs. We also observed that neuregulin-1 stimulation rapidly increased Dlg1 and PTEN (Fig. 2E), which resulted in more Dlg1-PTEN complexes (Fig. 2F).

Although neuregulin-1 activates PI3K-AKT and mitogen-activated protein kinase (MAPK) pathways in SCs (15, 20), inhibiting these pathways did not prevent the increase in Dlg1 and PTEN upon stimulation (Fig. 2G). This rapid increase was also not blocked by cycloheximide (Fig. 2G), which suggested that neuregulin-1 stimulation prevented Dlg1 and PTEN degradation. Indeed, blocking the SC proteasome increased Dlg1 and PTEN levels (fig. S5D). These proteins, and also AKT, were ubiquitinated in nonstimulated SCs and blocking the proteasome increased the amount of ubiquitinated proteins (fig. S5, E and F). Neuregulin-1 stimulation of these cells reduced the levels of ubiquitinated Dlg1 and PTEN (fig. S5E) but not of AKT (fig. S5F), which indicated that neuregulin-1 specifically prevents Dlg1 and PTEN ubiquitination and degradation, which leads to more Dlg1-PTEN complexes.

In mouse sciatic nerves, axonal neuregulin-1 type III stimulates myelination and the myelination rate [illustrated by the levels of myelin protein zero (P0) mRNA], and AKT activation (phospho-AKT levels) increases during the two first postnatal weeks (Fig. 3A and fig. S6). Dlg1 and PTEN levels also increased (Fig. 3A), whereas levels of polyubiquitinated Dlg1 and PTEN decreased (fig. S7), which suggests that neuregulin-1 stimulation also stabilizes Dlg1 and PTEN during myelination.

Although murine myelination continues for several months, the myelination rate and AKT activation decrease around 2 weeks after birth (Fig. 3A) (5). This inhibition of myelination correlates with the plateau reached by Dlg1 and PTEN expression (Fig. 3A), consistent with the proposed

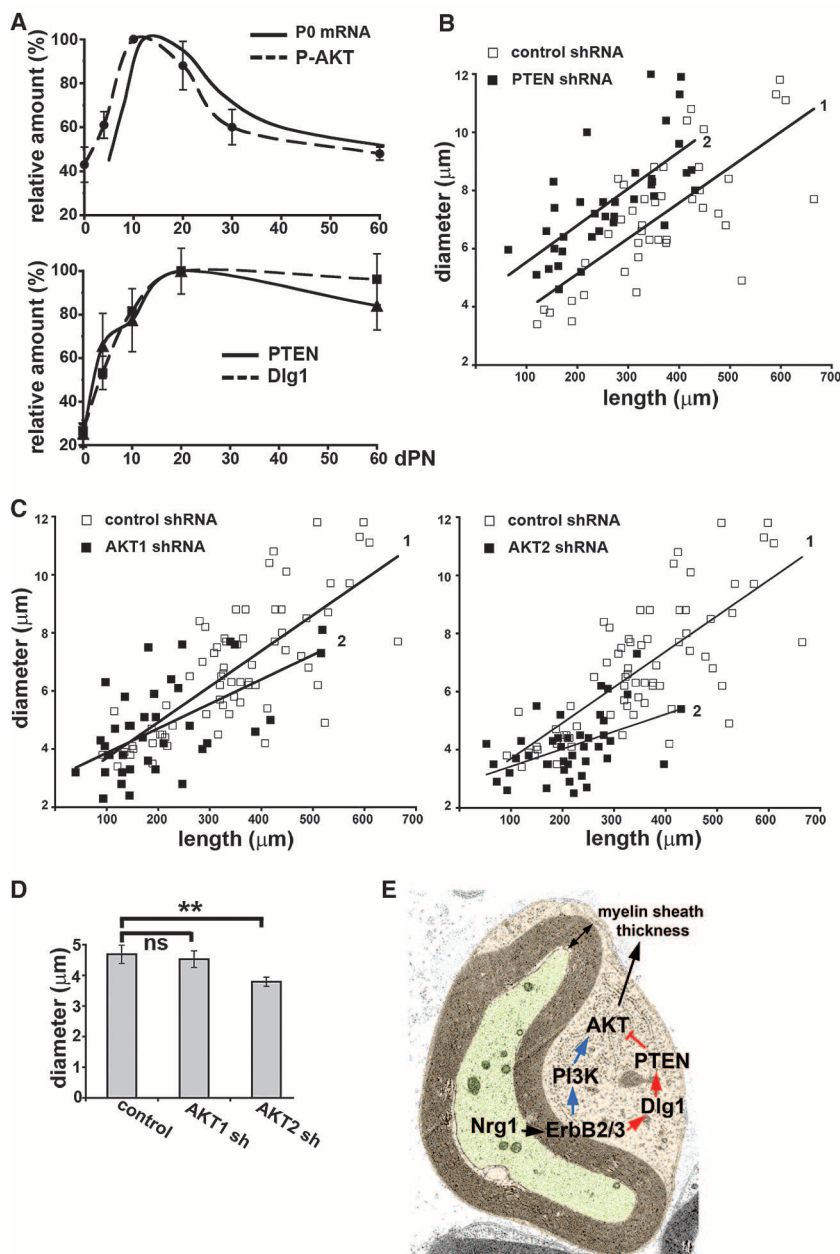


Fig. 3. PTEN, AKT1, and AKT2 participate in the regulation of myelin sheath thickness. **(A)** Relative amounts ($\pm$ SEM) of Dlg1, PTEN, and phospho-AKT analyzed during myelination in mouse sciatic nerve for at least three mice. (For original data, see fig. S6.) P0 mRNA data are derived from (5). Maximal expressions used as reference. Lines: Extrapolated regression curves. **(B)** Diameters versus lengths of SCs infected with PTEN or control shRNAs viruses. Lines: Linear regressions for control ($n = 42$) line 1 and PTEN ($n = 37$) line 2 shRNAs. **(C)** Diameters versus lengths of SCs infected with Akt1, Akt2, or control shRNAs viruses. Lines: Linear regressions for values of control (line 1) and Akt1 or Akt2 (line 2) shRNAs. $n = 39$ (AKT1), 41 (AKT2), and 69 (control). **(D)** Mean cell diameter ($\pm$ SEM), comparing silenced and nonsilenced SCs of the same length class ($<300 \mu\text{m}$) for $n = 31$ (AKT1), 34 (AKT2), and 22 (control). **(E)** Neuregulin-1 type III (Nrg1) on axons (green) stimulates erbB2 and erbB3 receptors in Schwann cells (brown). Receptor activation stimulates the PI3K-AKT pathway (blue arrows) promoting myelination. Neuregulin-1 stimulation also stabilizes Dlg1, which interacts with and stabilizes PTEN (red arrows). More PTEN is then available to inhibit AKT activation, which results in curbing of myelination.

inhibitory role of Dlg1 during myelination. To check the functional role of PTEN, we silenced its expression in SCs of sciatic nerves. As expected, reduction of PTEN expression correlated with an increase of phospho-AKT in infected cells (fig. S8). Myelinating SCs silenced for PTEN showed a thicker myelin sheath compared with control cells of similar length (Fig. 3B; $7.6 \pm 0.3 \mu\text{m}$ versus $6.1 \pm 0.3 \mu\text{m}$; length $< 405 \mu\text{m}$; $P < 0.01$). We conclude that PTEN works in concert with Dlg1 as myelination inhibitor.

Transgenic expression of constitutively active AKT in SCs by using the proteolipid protein promoter has been reported not to affect myelination (21). However, low expression levels or compensatory mechanism may have overshadowed an effect. Thus, using our experimental setting, we tested whether AKT, a target of Dlg1 and PTEN, is required for SC myelination. Silencing of the more ubiquitously expressed isoforms Akt1 or Akt2 (22) (fig. S9) resulted in shorter myelin sheaths (Fig. 3C), which showed that these proteins are required for myelination and/or myelin sheath maintenance. Cells silenced for Akt2 had significantly reduced diameters compared with control cells of similar lengths, whereas Akt1-silenced cells did not (Fig. 3D), which suggests that Akt2 is also required for correct myelin sheath thickness.

Taken together, our data indicate that a molecular mechanism involving the critical functions of Dlg1 and PTEN prevents SC overmyelination when they are stimulated with neuregulin-1 (Fig. 3E). Myelination is also regulated, however, by the

amount of neuregulin-1 expressed in axons (3, 4) and the availability of erbB2 and erbB3 receptors on SCs (23, 24). To test whether changes in these parameters are primarily responsible for curbing the myelination rate, we silenced Dlg1 in myelinating SCs at 20 dPN, after major myelination is stalling. Three weeks later, SCs silenced for Dlg1 had produced significantly larger myelin sheaths (Fig. 4A; $8.5 \pm 0.3 \mu\text{m}$ versus $5.3 \pm 0.2 \mu\text{m}$; $P < 0.001$), which showed that mechanisms stimulating SC myelination were still active at the time myelination was curbed. Thus, the Dlg1 brake appears to be the main mechanism reducing myelination at this age. When overmyelination was allowed to proceed for longer (6 weeks after injection), demyelination features were abundant, and only the shortest myelinated cells remained (fig. S10). Thus, the cells with the most oversized myelin are unstable, consistent with observations in several peripheral neuropathies (7).

Dlg1 interacts with the CMT4B-disease protein MTMR2 (8). We found that some Dlg1-silenced cells displayed small outfoldings (Fig. 4B) that occasionally expanded in large focally folded myelin sheaths, reminiscent of structural defects in CMT4B (Fig. 4C). This suggested that Dlg1 function might be altered in MTMR2 mutant SCs. Therefore, we silenced Dlg1 in SCs of the sciatic nerve of MTMR2 mutant mice. Dlg1-silenced cells were morphologically equal to control cells (Fig. 4D, diameter: $6.4 \pm 0.4 \mu\text{m}$ versus $6.7 \pm 0.3 \mu\text{m}$, $P = 0.47$), showing that Dlg1 was not inhibiting myelination in these cells. We pro-

pose that the myelination brake triggered by Dlg1 is missing in CMT4B-diseased SCs, which results in overmyelination, defects in myelin deposition, and, finally, demyelination.

Over 50 years ago, Rushton showed in a pivotal paper (25) that myelin sheath diameter is optimally sized toward a certain NCV. The myelination brake involving Dlg1 and PTEN appears to be essential for this optimization, in addition to axonal stimulation.

Perturbation of this basic regulatory mechanism provides a mechanistic basis for the etiology of hereditary peripheral neuropathies characterized by hypermyelination, and the control of enzymatic activities involved in this mechanism, PTEN and AKT, may serve as therapeutic strategies for these diseases.

References and Notes

1. S. G. Waxman, *Muscle Nerve* **3**, 141 (1980).
2. S. G. Waxman, *Curr. Biol.* **7**, R406 (1997).
3. G. V. Mikhailov et al., *Science* **304**, 700 (2004).
4. C. Taveggia et al., *Neuron* **47**, 681 (2005).
5. N. Stahl, J. Harry, B. Popko, *Brain Res. Mol. Brain Res.* **8**, 209 (1990).
6. S. Bonneick et al., *Hum. Mol. Genet.* **14**, 3685 (2005).
7. U. Suter, S. S. Scherer, *Nat. Rev. Neurosci.* **4**, 714 (2003).
8. A. Bolis et al., *J. Neurosci.* **25**, 8567 (2005).
9. T. Yamanaoka, S. Ohno, *Front. Biosci.* **13**, 6693 (2008).
10. Materials and methods are available as supporting material on Science Online.
11. A. Bolis et al., *J. Neurosci.* **29**, 8858 (2009).
12. D. M. Pegtel et al., *Curr. Biol.* **17**, 1623 (2007).
13. J. R. Chan et al., *Science* **314**, 832 (2006).
14. M. Özgüç et al., *J. Neurosci.* **30**, 4120 (2010).
15. K. A. Nave, J. L. Salzer, *Curr. Opin. Neurobiol.* **16**, 492 (2006).
16. N. B. Adey et al., *Cancer Res.* **60**, 35 (2000).
17. A. Contreras-Paredes, E. De la Cruz-Hernández, I. Martínez-Ramírez, A. Dueñas-González, M. Lizano, *Virology* **383**, 78 (2009).
18. L. Salmena, A. Carracedo, P. P. Pandolfi, *Cell* **133**, 403 (2008).
19. M. Valiente et al., *J. Biol. Chem.* **280**, 28936 (2005).
20. T. Ogata, S. Yamamoto, K. Nakamura, S. Tanaka, *Mol. Neurobiol.* **33**, 51 (2006).
21. A. I. Flores et al., *J. Neurosci.* **28**, 7174 (2008).
22. M. Hanada, J. Feng, B. A. Hemmings, *Biochim. Biophys. Acta* **1697**, 3 (2004).
23. A. N. Garratt, O. Voiculescu, P. Topilko, P. Charnay, C. Birchmeier, *J. Cell Biol.* **148**, 1035 (2000).
24. S. Chen et al., *J. Neurosci.* **26**, 3079 (2006).
25. W. A. Rushton, *J. Physiol.* **115**, 101 (1951).
26. We thank J. Salzer for his help with cocultures, N. Mantei, C. Perrin-Tricaud for critically reading the manuscript, and the Light Microscopy Center and Electron Microscopy Center at ETH facilities of the ETHZ for technical help. Supported by the Swiss National Science Foundation (N.T. and U.S.), the National Centre of Competence in Research "Neural Plasticity and Repair," and ETH Zürich (ETHIRA).

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1187735/DC1
Materials and Methods
Figs. S1 to S10
References

1 February 2010; accepted 7 April 2010

Published online 6 May 2010;

10.1126/science.1187735

Include this information when citing this paper.

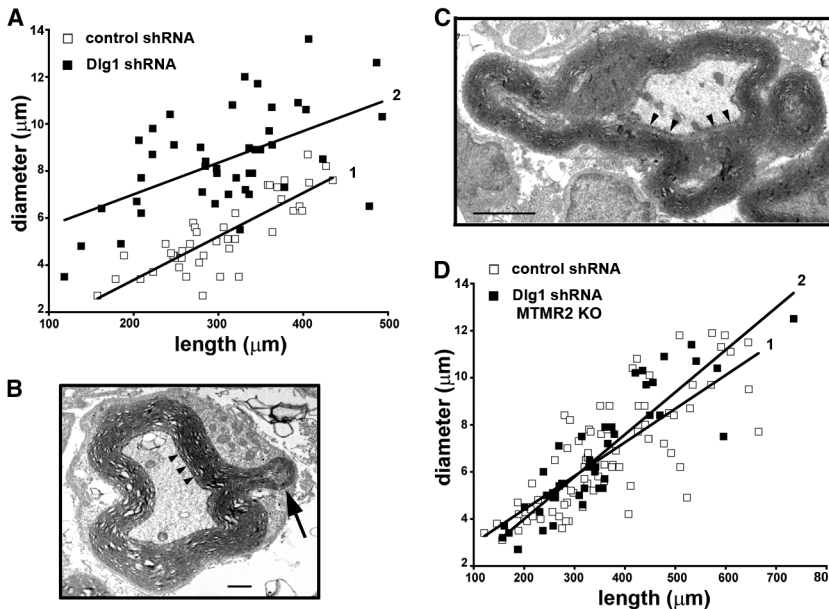


Fig. 4. Dlg1 is essential to curb myelination. (A) Dlg1 or control shRNAs adenoviruses were injected in mice after myelination slowdown (20 dPN). (Graph) Diameters versus lengths of infected SCs 3 weeks after injection. Lines: Linear regressions for values of control ($n = 41$) (line 1) and Dlg1 ($n = 43$) (line 2) shRNAs. (B) Hypermyelinated fiber expressing Dlg1 shRNA and PLAP (arrowheads) displays a typical outfolded (arrow). Scale bar, 800 nm. (C) Myelinating SC expressing Dlg1 shRNA and PLAP (arrowheads) show excessive and aberrant myelin deposition with large outfoldings. Scale bar, 2 μm . (D) Dlg1 or control shRNAs viruses were injected in sciatic nerves of MTMR2 mutant pups. (Graph) Diameters versus lengths of infected SCs. Lines: Linear regressions for values of control ($n = 80$) (line 1) and Dlg1 ($n = 45$) (line 2) shRNAs viruses.

NEW PRODUCTS FOCUS: SAMPLE HANDLING/STORAGE

RACK RUNNER ROBOT

The Rack Runner is a fast, flexible robotic tool that can integrate multiple -20°C and -80°C Sample Access Manager (SAM) storage systems into a cluster or link storage with the Hamilton liquid handling workstation. The Rack Runner's range enables scalable integration of multiple Hamilton and third-party systems, including incubators, readers, washers, and shakers. The Rack Runner is designed to meet the evolving storage needs of customers in biobanking, virus banking, DNA and RNA research, and pharmaceutical discovery and development. Users can start with a single SAM, adding and linking additional systems as capacity requirements grow in the future. The small and agile Rack Runner moves on a linear drive that is extendable from 80 centimeters to 8 meters. With wireless control, the robot moves at up to 2.5 meters per second.



Hamilton Storage Technologies

For info: 508-544-7000 | www.hamilton-storage.com

TUBE-TO-PLATE LIQUID HANDLER

The Sciclone G3 liquid handler, with the ability to transfer 24 samples at once, features innovations including a novel rack system, simplified decapping and tube loading, on-deck sensors, and an integrated bar code reader to identify and track samples. Other innovations include pioneering tube-to-plate reformatting technology, noncontact liquid level and clog detection, advanced blood tube management, and easy-to-use yet flexible software. The instrument provides a simple and fast method for tube-to-plate transfer. The Sciclone G3 can easily manage and process 96 blood tubes. Its on-deck sensors and integrated bar code reader identify and track samples, providing a comprehensive and direct approach to sample handling. The liquid level sensor and tube management capabilities allow for processing samples directly from blood tubes.

Caliper Life Sciences

For info: 508-435-9500 | www.caliperLS.com

THERMAL CYCLERS

Designed to save time and energy, the TC-Plus thermal cyclers and satellite units feature a stackable configuration to reduce the bench space required and facilitate networking. They feature the Thermal Energy Recovery System, which saves energy by harnessing the heat vented during the cooling phase for use in the next heating phase. With a ramp rate of 5°C per second, the aluminum block is one of the fastest available. These thermal cyclers also feature a large color touchscreen for quick program setup. At least a thousand programs can be stored and the front-positioned USB port makes it easy to transfer protocols between cyclers and laboratories. A front-loading sample drawer makes the units stackable, which, together with the front-to-back venting and airflow, enables close side-by-side positioning, minimizing the total footprint.

Bibby Scientific

For info: +44-(0)-1785-812121 | www.bibbyscientific.com

SAMPLE MANAGEMENT SYSTEM

The Track-IT is a sample management system interfaced to a two-dimensional code reader that allows the user to reliably and efficiently organize and track individual samples. Track-IT is a complete sample management software package that is suitable for biobanks and biorepositories as well as molecular biology, cellular biology, clinical, chemical, biochemistry, immunology, high throughput screening, and in vitro fertilization laboratories. Fully compatible with most commercial barcode readers and scanners, Track-IT allows users to quickly catalog and organize samples. The central information database produced by Track-IT is customizable to match the workflow of any laboratory, and to match the unique aspects of different sample types.

Micronic Europe

For info: +31-320-277070 | www.micronic.com/track-it/htm

UV STERILIZATION CABINET

The GLE-UVSC ultraviolet (UV) sterilization cabinet is designed for molecular biology laboratories carrying out polymerase chain reaction (PCR) processes. The cabinet provides decontamination of reagents and equipment before carrying out sensitive PCR reactions, which is particularly important when amplifying DNA sequences that are in limited supply or of low copy number. Four timer-controlled, 15-watt UV bulbs enable the user to control the exposure time and dose of high-energy UV radiation used to denature the nucleic acids that cause unwanted background contamination. A timer override switch allows the UV source to remain on indefinitely. Safety interlocks turn off the bulbs when the doors are opened.

Scie-Plas

For info: +44-(0)-1223-427888 | www.scie-plas.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.